

STIC-Biotech/ChemLib

73796

From: Chan, Christina
Sent: Tuesday, August 20, 2002 3:19 PM
To: Wilder, Cynthia; STIC-Biotech/ChemLib
Subject: RE: Rush sequence search 09/835,082

12 E11

Please rush. Thanks Chris

-----Original Message-----

From: Wilder, Cynthia
Sent: Tuesday, August 20, 2002 12:14 PM
To: Chan, Christina
Subject: Rush sequence search 09/835,082

Ms. Chan,

I am requesting a Rush sequence search for 09/835,082 due to a low docket. Please forward your approval to STIC.

Please search the nucleic acid encoded by SEQ ID NO: 1 and SEQ ID NO: 3.

Thank you!

Cynthia B. Wilder, Ph.D.

Art Unit 1637

703/305-1680

Office No. 12 E11

Mailbox No. 10 E12

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 8/20/02
Date Completed: 8/21/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: 2
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 05
WWW/Internet: _____
Other (specify): _____

[The body of the page contains extremely faint, illegible text, likely bleed-through from the reverse side. The text is organized into several paragraphs, but the characters and words are too light to be transcribed accurately.]

892	00Y	aatgctcttgagaaacatggaattggtgtttctctttgggatcaaatggtctagaataatt	951
893	00Y		
894	00Y		
895	00Y		
896	00Y		
897	00Y		
898	00Y		
899	00Y		
900	00Y		
901	00Y		
902	00Y		
903	00Y		
904	00Y		
905	00Y		
906	00Y		
907	00Y		
908	00Y		
909	00Y		
910	00Y		
911	00Y		
912	00Y		
913	00Y		
914	00Y		
915	00Y		
916	00Y		
917	00Y		
918	00Y		
919	00Y		
920	00Y		
921	00Y		
922	00Y		
923	00Y		
924	00Y		
925	00Y		
926	00Y		
927	00Y		
928	00Y		
929	00Y		
930	00Y		
931	00Y		
932	00Y		
933	00Y		
934	00Y		
935	00Y		
936	00Y		
937	00Y		
938	00Y		
939	00Y		
940	00Y		
941	00Y		
942	00Y		
943	00Y		
944	00Y		
945	00Y		
946	00Y		
947	00Y		
948	00Y		
949	00Y		
950	00Y		
951	00Y		
952	00Y	ccagagaagaagctatgccaattgctgatgctttgggcacaaatccctcagacagctctg	1011
953	00Y		
954	00Y		
955	00Y		
956	00Y		
957	00Y		
958	00Y		
959	00Y		
960	00Y		
961	00Y		
962	00Y		
963	00Y		
964	00Y		
965	00Y		
966	00Y		
967	00Y		
968	00Y		
969	00Y		
970	00Y		
971	00Y		
972	00Y		
973	00Y		
974	00Y		
975	00Y		
976	00Y		
977	00Y		
978	00Y		
979	00Y		
980	00Y		
981	00Y		
982	00Y		
983	00Y		
984	00Y		
985	00Y		
986	00Y		
987	00Y		
988	00Y		
989	00Y		
990	00Y		
991	00Y		
992	00Y		
993	00Y		
994	00Y		
995	00Y		
996	00Y		
997	00Y		
998	00Y		
999	00Y		
1000	00Y		
1001	00Y		
1002	00Y		
1003	00Y		
1004	00Y		
1005	00Y		
1006	00Y		
1007	00Y		
1008	00Y		
1009	00Y		
1010	00Y		
1011	00Y		
1012	00Y	tggcgggtcacctggaacccgcacatcgaaattctgcaaacacacgatacttgttaagtgg	1071
1013	00Y		
1014	00Y		
1015	00Y		
1016	00Y		
1017	00Y		
1018	00Y		
1019	00Y		
1020</			

Db 601 ccatgattcttctgttgggaaagaaatgatgctatgaaattggtggtgtatt 660
 Qy 2211 tgagaagataatcattgcttatgcaaatgagctgaattgataaaaaacccaaatata 2270
 Db 661 tgagaagataatcattgcttatgcaaatgagctgaattgataaaaaacccaaatata 720
 Qy 2271 gctatgaagtgcgtggcaagtttacttttttctgagtcttactacaact 2320
 Db 721 gctatgaagtgcgtggcaagtttacttttttctgagtcttactacaact 770
 RESULT 11
 AA245116
 ID AA245116 standard; DNA; 930 BP.
 XX
 AC AA245116;
 XX
 DT 28-FEB-2020 (first entry)
 XX
 DE UDP-glucuronosyltransferase 1 (UGT1) exon 1H nucleotide sequence.
 XX
 KW Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
 KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
 KW unconjugated hyperbilirubinemia; drug metabolism; transgenic animal;
 KW pharmacogenetic screening; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0957322-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 04-MAY-1999; 99MO-US09702.
 XX
 PR 07-MAY-1998; 98US-0084807.
 XX
 PA (AXYS-) AXYS PHARM INC.
 XX
 PI Penny L, Galvin M;
 XX
 WPI: 2000-052981/04.
 DR P-PSDB; AAY57098.
 XX
 PT New nucleic acid representing polymorphisms in the human uridine
 PT diphosphate glucuronosyltransferase gene, used for diagnosis and
 PT evaluation of drug metabolism -
 XX
 PS Examples; Page 39-41; 63pp; English.
 XX
 CC Nucleotide sequences AA245110-245118 are exons 1A-1J of the human
 CC uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
 CC family of enzymes that catalyse the glucuronic acid conjugation of a wide
 CC range of endogenous and exogenous substrates including phenols,
 CC alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
 CC result in toxic substances being converted to compounds which are more
 CC water soluble and are excreted. The invention relates to and identifies
 CC UGT1 polymorphisms (AA245004-245041). The polymorphism sequences are
 CC useful as probes for detecting UGT1 locus polymorphisms, indicative of
 CC altered UGT1 expression or activity. These polymorphisms are associated
 CC with Crigler-Najjar and Gilbert syndromes (unconjugated
 CC hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
 CC is used to predict the rate of metabolism of UGT1 substrates, possible
 CC drug-drug interactions and adverse side effects (i.e. to optimize drug
 CC dosage), and to screen for diseases caused by exposure to toxins and to
 CC study the effects of polymorphisms on enzymatic activity. The UGT1
 CC sequences, including polymorphisms, can also be used to produce the
 CC corresponding protein (or its fragments) or to generate transgenic
 CC animals or modified cells e.g. for pharmacogenetic screening.
 XX
 SQ Sequence 930 BP; 228 A; 205 C; 212 G; 285 T; 0 other;

Query Match 33.0%; Score 766.6; DB 21; Length 930;
 Best Local Similarity 91.6%; Pred. No. 2.1e-207;
 Matches 811; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
 Qy 19 atggcttgcaacaggggtgacacagcccccttctctatgtgtgtctgtctgacctgt 78
 Db 1 atggcttgcaacaggggtgacacagcccccttctctatgtgtgtctgtctgacctgt 60
 Qy 79 ggccttccagagcaggggaagctactgtgtggtgcccaggtgagccactggttcacc 138
 Db 61 ggccttccagagcaggggaagctactgtgtggtgcccaggtgagccactggttcacc 120
 Qy 139 atgaggtcggtggtgagaaactcattctcaggggagcagctgaggtggttagtcagcca 198
 Db 121 atgaggtcggtggtgagaaactcattctcaggggagcagctgaggtggttagtcagcca 180
 Qy 199 gaggtgaggtggcaactgggaagatcaactgaattgcacagtgaaagacttattcaactca 258
 Db 181 gaggtgaggtggcaactgggaagatcaactgaattgcacagtgaaagacttattcaactca 240
 Qy 259 tataccttgagagatctgacccggaggttcaagccttttgcctatgctcaatggaagca 318
 Db 241 tataccttgagagatctgacccggaggttcaagccttttgcctatgctcaatggaagca 300
 Qy 319 caagtacgaagtataattctctatgttggtggttcaatgacattttgtgacttatt 378
 Db 301 caagtacgaagtgtgtttctctattcttctgagttcactcaatggttttttaacttatt 360
 Qy 379 ttttcaaatgacagaggttgttttaagacaaaataatagtagaatacttaagagagagt 438
 Db 361 ttttgcagctgacagaggttgttttaagacaaaataatagtagaatacttaagagagagt 420
 Qy 439 tcttttgatgacaggttctctgacatcttttgataactgtggttaattgtgccaatat 498
 Db 421 tcttttgatgacaggttctctgacatcttttgataactgtggttaattgtgccaatat 480
 Qy 499 ttctccctccctccgtggttctgcccaggggaatactttgccactatcttgaaagagt 558
 Db 481 ttctccctccctccgtggttctgcccaggggaatactttgccactatcttgaaagagt 540
 Qy 559 gcacagtgccctgctctctctctctatgtcccagaaattctcttaggggttctcagatgcc 618
 Db 541 gcacagtgccctgctctctctctctatgtcccagaaattctcttaggggttctcagatgcc 600
 Qy 619 atgactttcaagagagagatgacggaacacacatcatctgacacttgaggagacatttattatgc 678
 Db 601 atgactttcaagagagagatgacggaacacacatcatctgacacttgaggagacatttattatgc 660
 Qy 679 caccgttttttcaaaaatgcccctagaaatagcctctgaaattctccaaacacctgttaccg 738
 Db 661 caccgttttttcaaaaatgcccctagaaatagcctctgaaattctccaaacacctgttaccg 720
 Qy 739 gagtatgactctacagccacacacatcaatttgggtgtgtgcgaacggactttgttttggac 798
 Db 721 gagtatgactctacagccacacacatcaatttgggtgtgtgcgaacggactttgttttggac 780
 Qy 799 tatcccaaacccgtgagtcgcaacatgactctcattggtggtatcaactgccaatcagaga 858
 Db 781 tatcccaaacccgtgagtcgcaacatgactctcattggtggtatcaactgccaatcagaga 840
 Qy 859 aagcgtgtgctatggaatttgaagcctacattaatgcttcttga 903
 Db 841 aagcgtgtgctatggaatttgaagcctacattctctctttagcaccattagga 885

RESULT 12
 AA245115
 ID AA245115 standard; DNA; 951 BP.
 XX
 AC AA245115;
 XX
 DT 28-FEB-2000 (first entry)
 XX


```

Db 541 gacgtgcagcaggctggcattctagattactttcttacttgaacatggcctgttg 600
Qy 1972 ggaagtgcggattcaagggtggtccacacgctgccctactgcacatggcagttttaac 1981
Db 601 ggaagtgcggattcaagggtggtccacacgctgccctactgcacatggcagttttaac 660
Qy 1982 ttatctttggtcttcagatggttgcaattgactcttaaccaataatggtcagtcctc 2041
Db 661 ttatctttggtcttcagatggttgcaattgactcttaaccaataatggtcagtcctc 720
Qy 2042 atctctgtctcttctatagtgccacactgtgtgtttaagaagggaagctttgtacct 2101
Db 721 atctctgtctcttctatagtgccacactgtgtgtttaagaagggaagctttgtacct 780
Qy 2102 ttagaagtaggtggaatgaatgaatggtctggtggtgacactgagacagcgtatgattc 2161
Db 781 ttagaagtaggtggaatgaatgaatggtctggtggtgacactgagacagcgtatgattc 840
Qy 2162 ttgcttgggggaaaaaagaatgactatgaaatggtggtggtggttattgagaagataa 2221
Db 841 ttgcttgggggaaaaaagaatgactatgaaatggtggtggtggttattgagaagataa 900
Qy 2222 tcattcttatgtcaaatgagctgaattgtataaaaccccaataacacgtatgaagt 2281
Db 901 tcattcttatgtcaaatgagctgaattgtataaaaccccaataacacgtatgaagt 960
Qy 2282 ctggggcaagtttactttttttctgattgtttctctacaact 2320
Db 961 ctggggcaagtttactttttttctgattgtttctctacaact 999

```

RESULT 8

```

AAH51404
ID AAH51404 standard; DNA; 1001 BP.
XX AC AAH51404;
XX DT 29-AUG-2001 (first entry)
XX DE Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 295.
XX KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucuronosyl transferase; UGT2; asthma; hepatotoxicity;
XX KW zileuton; ds.
OS Homo sapiens.
PN W0200058508-A2.
XX PD 05-OCT-2000.
XX PF 24-MAR-2000; 2000WO-IB00403.
XX PR 25-MAR-1999; 99US-0126269.
XX PR 30-APR-1999; 99US-0131961.
XX PA (GEST ) GENSET.
XX PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX DR WPI; 2000-638353/61.
XX PR Polynucleotides comprising sequences from malate decarboxylase
XX enzyme-related biallelic markers used for genotyping -
PS Claim 13; Page 487-488; 673pp; English.
XX

```

```

CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucuronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAH52905-AAH62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC a population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatotoxicity on treatment with
CC zileuton.
XX
SQ Sequence 1001 BP; 293 A; 171 C; 210 G; 326 T; 1 other;

```

Query Match 37.5%; Score 870.6; DB 21; Length 1001;

Best Local Similarity 99.9%; Pred. No. 5.6e-237;

Matches 870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1450 ctcaactgtgtaccacgtaccattcttgacgtgattgttctctctgtgctgctgctg 1509
Db 1 ctcaactgtgtaccacgtaccattcttgacgtgattgttctctctgtgctgctgctg 60
Qy 1510 acagtgtgctctacacctttaaattgtgtgtcttattgctacccggaatgcttgggaaa 1569
Db 61 acagtgtgctctacacctttaaattgtgtgtcttattgctacccggaatgcttgggaaa 120
Qy 1570 aaagggcagtgtaagaagcccaaaatcccaagccattgagaagtggtgggaaataa 1629
Db 121 aaagggcagtgtaagaagcccaaaatcccaagccattgagaagtggtgggaaataa 180
Qy 1630 ggtaaaatttgaaccattcccttagtctcattccaaactgaaacacagacagtggttaa 1689
Db 181 ggtaaaatttgaaccattcccttagtctcattccaaactgaaacacagacagtggttaa 240
Qy 1690 ttcatattattcttatttaaggaaatacttttgataaataatcagccacagtgcttta 1749
Db 241 ttcatattattcttatttaaggaaatacttttgataaataatcagccacagtgcttta 300
Qy 1750 aaaaattctcttaaaaaataaataaactcgtcgtcagtaagaatatttgaatgta 1809
Db 301 aaaaattctcttaaaaaataaataaactcgtcgtcagtaagaatatttgaatgta 360
Qy 1810 tcgtgccccctccggtgtgttttgatcagatgacatgtgccatttttcagagacgtgca 1869
Db 361 tcgtgccccctccggtgtgttttgatcagatgacatgtgccatttttcagagacgtgca 420
Qy 1870 gacagctggcattctagattactttcttactctgaacacatgctgtttggagtcg 1929
Db 421 gacagctggcattctagattactttcttactctgaacacatgctgtttggagtcg 480
Qy 1930 ggaattcaagtggtccacacgctgccccctactgcaaatggcagtttttaattcttctt 1989
Db 481 ggaattcaagtggtccacacgctgccccctactgcaaatggcagtttttaattcttctt 540
Qy 1990 tggcttcgcagatgttgcaattgattccttaaccataatggtcagtcctcctctct 2049

```


to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1655 BP; 402 A; 388 C; 403 G; 462 T; 0 other;

Query Match 66.9%; Score 1551.8; DB 23; Length 1655;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY	1	ctcagctgcagttctctctgtaggcttgcacagagtgagaccagccctctctctatgtgtg	60
DB	45	ctcagctgcagttctctgtaggcttgcacagagtgagaccagccctctctctatgtgtg	104
QY	61	tgtctgctgtgacctgtgaccttgccaggcagggagagctactgtgtgctccatggat	120
DB	105	tgtctgctgtgacctgtgaccttgccaggcagggagagctactgtgtgctccatggat	164
QY	121	ggagccactgggttcacacatgaggttcggttgtagaataactctctcaggggcacatgag	180
DB	165	ggagccactgggttcacacatgaggttcggttgtagaataactctctcaggggcacatgag	224
QY	191	gtggtgtgtagctgcagaggtgagttggcaactgggaagatcactgaattgcacagtg	240
DB	225	gtggtgtgtagctgcagaggtgagttggcaactgggaagatcactgaattgcacagtg	284
QY	241	aagacttatcaactctatataccttgagagatctggaccggaggtcgaagcttttggcc	300
DB	285	aagacttatcaactctatataccttgagagatctggaccggaggtcgaagcttttggcc	344
QY	301	catgtcgaatggaaagcaccaagtacgaagtatatattctctatttaattgggttcatacaat	360
DB	345	catgtcgaatggaaagcaccaagtacgaagttgtttctctatttctgagttccatccaat	404
QY	361	gacattttgacctattttttccaaattgcagagagttgttttaaagacaaaataatagta	420
DB	405	ggtttttttaactatttttttcgactgcagagagttgttttaaagacaaaataatagta	464
QY	421	gaatactaaaggagaggtctttttgatgcagtggtttctcgatctcttttgataactgtggc	480
DB	465	gaatactaaaggagaggtctttttgatgcggtgtgtttcttgatctcttttgatgcgtggcg	524
QY	481	ttaattgttgccaaatatttctccctcccttcgtggtgttttcggcaggggaataactttgc	540
DB	525	ttaattgttgccaaatatttctccctcccttcgtggtgttttcggcaggggaataactttgc	584
QY	541	cactatcttgaagaaggtgcacagtgccctgcctctcttctctatgtcccagaaattctc	600
DB	585	cactatcttgaagaaggtgcacagtgccctgcctctcttctctatgtcccagaaattctc	644
QY	601	ttaggtgtctcagatgccatgactttcaagagagagtagcagaaaccacatactgcaacttg	660
DB	645	ttaggtgtctcagatgccatgactttcaagagagagtagcagaaaccacatactgcaacttg	704
QY	661	gaggaaactatttatcgccacaggttttttcaaaaatgcctcagaatactgacctgaaatt	720
DB	705	gaggaaactatttatcgccacaggttttttcaaaaatgcctcagaatactgacctgaaatt	764
QY	721	ctcaaaacacctgttacgggagtagtatctctcagcgccacacatacaattttggtgttgca	780


```

Db 1870 gtgcagacagctggcattctagattacttttcttactctgaaacatggcgtgtttggga 1929
QY 1925 gtgcggattcaaaagtggtccacccgtgcccactgcccactgcaaatggcagttttaactcta 1984
Db 1930 gtgcggattcaaaagtggtccacccgtgcccactgcccactgcaaatggcagttttaactcta 1989
QY 1985 tcttcttgctctgcagatggttgcaattgatcctttaaccataatggcagtcctcactc 2044
Db 1990 tcttcttgctctgcagatggttgcaattgatcctttaaccataatggcagtcctcactc 2049
QY 2045 tctg-cctgcttctaatggtgcacc-ttgtgtgttttaagaagggaagctttgtacctt 2103
Db 2050 tctg-cctgcttctaatggtgcacc-ttgtgtgttttaagaagggaagctttgtacctt 2109
QY 2104 agag-gtagtgtaaatgaatgcttggtgagtgcaactgcaactgagacacatatgatttctt 2163
Db 2110 agag-gtagtgtaaatgaatgcttggtgagtgcaactgcaactgagacacatatgatttctt 2169
QY 2164 gctt-gggggaaaagaatgctgctatgaaattggtggtgtgtattttgagaagataatc 2223
Db 2170 gctt-gggggaaaagaatgctgctatgaaattggtggtgtgtattttgagaagataatc 2229
QY 2224 attgcttatgtcaaatgagctgaatttgataaaaccccaaaatacacgctatgaaagtct 2283
Db 2230 attgcttatgtcaaatgagctgaatttgataaaaccccaaaatacacgctatgaaagtct 2289
QY 2284 gggcnaagtttaactttttcttgatgttttctcaact 2320
Db 2290 gggcnaagtttaactttttcttgatgttttctcaact 2326

RESULT 4
AA027370
ID AA027370 standard; cDNA; 2368 BP.
XX AC AA027370;
XX DT 27-JAN-1993 (first entry)
XX DE HUG-Br2.
XX KW Billirubin: UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 30..801
XX FT polyA_signal 2347..2352
XX FT /*tag= a
XX FT /*tag= b
XX FT /*number= 1
XX FT polyA_signal 2355..2360
XX FT /*tag= c
XX FT /*number= 2
XX PN W09212987.A.
XX PD 06-AUG-1992.
XX PF 10-JAN-1992; 92WO-US00282.
XX PR 10-JAN-1991; 91US-0639453.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX PI Owens IS, Ritter JK;
XX XX
XX DR WPI: 1992-284593/34.
XX DR P-PSDB; AAR26154.
XX Isolated gene locus UGrl, DNA segments and diagnostic probes -
XX PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
XX

```

PT types I and II

XX Disclosure; Fig 9A-I; 99pp; English.

CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have been isolated. They are referred to as HUGBr1 (AAQ27369) and HUGBr2 (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and, upon expression individually in COS-1 cells, encode isoforms that catalyze the formation of the two bilirubin monoglucuronides and the diglucuronide.

CC The cDNAs contain identical 3' ends (1469 bp in length) to each other and to that of the human phenol transferase cDNA, HUGP1 (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).

CC In contrast, they have unique 5' ends.

XX Sequence 2368 BP; 609 A; 532 C; 566 G; 661 T; 0 other;

Query Match 68.0%; Score 1576.8; DB 13; Length 2368;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1847; Conservative 0; Mismatches 392; Indels 5; Gaps 3;

QY 81 ctttgcgagcagggagcactactgttagtgcccatggatggagccactggttcaactat 140

Db 101 ctgggctcgagctggaaagggtgtgtgggtgcccactgatggcagccctggtcagcat 160

QY 141 gagtcggtgtgtggaactcattctcaggggcatgaggtgtgttagtcacgcaga 200

Db 161 gggggaggccttgcggagctccatgccagagccaccaggcgtgtgctcaccaccaag 220

QY 201 ggtgagttgcgaactgggaagatcactgaattgcacagtgagacttattcaactcata 260

Db 221 aggtgaatatgcacatcaagaagagaaattttccacctgacacgctatgctgtccatg 280

QY 261 taccctggaggtatctggaccggagggttcaaggcttttggccatgctcaat--ggaagac 317

Db 281 gaccacagaaggaatttgatcggttacgtgggtacactcaagggttctttgaaacaga 340

QY 318 acaagtcaggaattatctcttatttaaggtgtcacaatgacatttttgacttatt 377

Db 341 acattcttgagagatatcttagaagtatggcaattatgacaattgtatctttggccct 400

QY 378 ttttcaaatgagggaggtttgtttaagacaaaaaattagataacttaaaagagag 437

Db 401 tcataggtgtgtgtgagctactgcataatgagccctgatcaggcactgaatgctac 460

QY 438 tcttttgatgcaggtttctcgatccttttgataactgtgcttaattgtgccaaata 497

Db 461 ttccttgatgtgtgttttaacagaccccgcttaacctctgcggggcggtgctgctaagta 520

QY 498 tttctccctccctccgtgtgttcttcgcaagggaataactttggccactatcttgaagaagg 557

Db 521 cctgtcattctctgtgttttttggagggtacattccatgtgacttagactttaagg 580

QY 558 tgcacagtcctgctcctctctctatgtcccagaattctcttaggttctcagatgc 617

Db 581 caacaggtccaaatcctctctctatactcctaagtactaacagacaattcagacca 640

QY 618 catgactttcaaggagagatgacggaaccacatcatgcaacttgaggagaacttattatg 677

Db 641 catgacattctcgaagggtcgaagaacatgctctaccctgtgcccctgctctacattg 700

QY 678 ccacgcttttttcaaaaatgcctctgaaatagcctctgaaattctccaaacactgttac 737

Db 701 ccatacttttttgcctcttatgcaagctctgctctgagctttttcagagagagtgac 760

QY 738 ggaatgatctctacagccacacatcaattgtgtgttgcaacgagactttgttttga 797

Db 761 tgcagggaactattgagctctgcatctgtctggctgttttagaagtgactttgtgaaga 820

QY 798 ctatcccaacccggtgatgcccaacatgattctatgtgtgtgtatcaactgccacagg 857

Db 821 ttaccctaggcccatcatgcccaatatggtcttcatgtggggggcaccataactgttgcacag 880

Qy 781 acggacttgggttggactatccccaccccgatgcccacacatgatcttcaattggtgt 840
 Db 781 acggacttgggttggactatccccaccccgatgcccacacatgatcttcaattggtgt 840
 Qy 841 atcaactgccatcagggaaagcgttgcctatggaatttgaagccctacatttaattgctct 900
 Db 841 atcaactgccatcagggaaagcgttgcctatggaatttgaagccctacatttaattgctct 900
 Qy 901 gagaaacatgaaattggtgtttctcttgggatacaatggtctcagaaattccagagaag 960
 Db 901 gagaaacatgaaattggtgtttctcttgggatacaatggtctcagaaattccagagaag 960
 Qy 961 aaagctatggcaattgctgatcttgggcaaaatccctcagacagctcctgtgcggtac 1020
 Db 961 aaagctatggcaattgctgatcttgggcaaaatccctcagacagctcctgtgcggtac 1020
 Qy 1021 actggaacccagacatcgaaatttggcaaacacacgatacttggtaagtggctaccccaa 1080
 Db 1021 actggaacccagacatcgaaatttggcaaacacacgatacttggtaagtggctaccccaa 1080
 Qy 1081 aacgatctggttggtaacccgatgaccggtgaccttattcacccatgctggttcccatggt 1140
 Db 1081 aacgatctggttggtaacccgatgaccggtgaccttattcacccatgctggttcccatggt 1140
 Qy 1141 gtttatgaagcatatgcaatggcgttcccatggtgatgacccctgttggatgacag 1200
 Db 1141 gtttatgaagcatatgcaatggcgttcccatggtgatgacccctgttggatgacag 1200
 Qy 1201 atggacaatgcaaaagcgcatgagactaaaggagctgagagtgaccctgaatgtttcgaa 1260
 Db 1201 atggacaatgcaaaagcgcatgagactaaaggagctgagagtgaccctgaatgtttcgaa 1260
 Qy 1261 atgaactctgaagatttagaaatgctctaaaaagcagtcataatgacaaaagtacaaag 1320
 Db 1261 atgaactctgaagatttagaaatgctctaaaaagcagtcataatgacaaaagtacaaag 1320
 Qy 1321 gagaacatcaatggcctctccagccttcacaaagcccggtggagccgctggacctg 1380
 Db 1321 gagaacatcaatggcctctccagccttcacaaagcccggtggagccgctggacctg 1380
 Qy 1381 gccgtgtctgggtgagttgtgatgagcaacaaaggcgccacacacctgcgcccga 1440
 Db 1381 gccgtgtctgggtgagttgtgatgagcaacaaaggcgccacacacctgcgcccga 1440
 Qy 1441 gccacgacctccacctggtaccagttaccattccttggacgtgattggtttcctctggcc 1500
 Db 1441 gccacgacctccacctggtaccagttaccattccttggacgtgattggtttcctctggcc 1500
 Qy 1501 gtcgtgtgacagtgcccttcacacatttaagtgtgtgttatggtcctaccggaatgc 1560
 Db 1501 gtcgtgtgacagtgcccttcacacatttaagtgtgtgttatggtcctaccggaatgc 1560
 Qy 1561 ttggggaaaaagggcagtttaagaaagcccaaatccaaagaccatttgagaagtgggt 1620
 Db 1561 ttggggaaaaagggcagtttaagaaagcccaaatccaaagaccatttgagaagtgggt 1620
 Qy 1621 gggaataaaggtaaaatttgaaacattcctcagtcatttcaaatctgaaaacagaaatc 1680
 Db 1621 gggaataaaggtaaaatttgaaacattcctcagtcatttcaaatctgaaaacagaaatc 1680
 Qy 1681 agtgttaattctatttatttatttaaggaatactttgtcataaataatcaatcagcccaag 1740
 Db 1681 agtgttaattctatttatttatttaaggaatactttgtcataaataatcaatcagcccaag 1740
 Qy 1741 agtgtttaaataattctttaaataataataatagactcgtcagtaagataattt 1800
 Db 1741 agtgtttaaataattctttaaataataataatagactcgtcagtaagataattt 1800
 Qy 1801 gaatatgatactgtgcccctccggtgtctttgatcagatgacatgtgccatttttcaga 1860
 Db 1801 gaatatgatactgtgcccctccggtgtctttgatcagatgacatgtgccatttttcaga 1860

Qy 1861 ggacgtgcagacagctggcattctagattacttttttactctgaaacatggcctgttt 1920
 Db 1861 ggacgtgcagacagctggcattctagattacttttttactctgaaacatggcctgttt 1920
 Qy 1921 ggagatgcggattcaaaagtgtgtcccaacgctgcacctactgcaaatggcagtttaatt 1980
 Db 1921 ggagatgcggattcaaaagtgtgtcccaacgctgcacctactgcaaatggcagtttaatt 1980
 Qy 1981 ctatcttttggcttctgcagatggttgcaattgacaccttaaccaataatgggtcagctct 2040
 Db 1981 ctatcttttggcttctgcagatggttgcaattgacaccttaaccaataatgggtcagctct 2040
 Qy 2041 catctctgctgctcatagtgccaccttgggttttaaaagaagggaagcctttgtacc 2100
 Db 2041 catctctgctgctcatagtgccaccttgggttttaaaagaagggaagcctttgtacc 2100
 Qy 2101 tttagagtgtagtgtaaatgaaatggctggctgagtcgactggaacagcatatgattt 2160
 Db 2101 tttagagtgtagtgtaaatgaaatggctggctgagtcgactggaacagcatatgattt 2160
 Qy 2161 cttgctttggggaaaaagatgactatgaaattgggtgggtgtatttggagaagata 2220
 Db 2161 cttgctttggggaaaaagatgactatgaaattgggtgggtgtatttggagaagata 2220
 Qy 2221 atcattgcttatgtcaaatggagctgaattttgataaaaccccaaaacacagctatgaagt 2280
 Db 2221 atcattgcttatgtcaaatggagctgaattttgataaaaccccaaaacacagctatgaagt 2280
 Qy 2281 gctgggcaaatgttacttttttttttttttttttttttttttttttttttttttttt 2320
 Db 2281 gctgggcaaatgttacttttttttttttttttttttttttttttttttttttttttt 2320

RESULT 3

AA027369
 ID AA027369 standard; cDNA; 2351 BP.
 XX
 AC AA027369;
 XX
 DT 27-JAN-1993 (first entry)
 XX
 DE HUG-Br1.
 XX
 KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
 monoglucuronide; diglucuronide; ss.
 XX
 OS Homo sapiens.
 XX

Key Location/Qualifiers
 CDS 16..784
 FT /*tag= a
 FT polyA_signal 2330..2335
 FT /*tag= b
 FT /*number= 1
 FT polyA_signal 2338..2343
 FT /*tag= c
 FT /*number= 2

XX W09212987-A.
 PN
 XX
 PD 06-AUG-1992.
 XX
 PF 10-JAN-1992; 92WO-US00282.
 XX
 PR 10-JAN-1991; 91US-0639453.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX
 PI Owens IS, Ritter JK;
 XX
 DR WPI; 1992-284593/34.
 DR P-PSDB; AAR26153.
 XX

XX WPI; 2002-075093/10.
 DR P-PSDB: ABB04429.
 XX Combinations of flavopiridol and an agent that increases conjugative
 PT enzyme activity or glucuronosyltransferase activity, with reduced side
 PT effects, for treating cancer -
 XX Claim 80; Page 137-140; 145pp; English.
 XX The present invention relates to a method of reducing the toxicity of
 CC flavopiridol by administration in combination with a second agent that
 CC increases conjugative enzyme activity or glucuronosyltransferase
 CC activity. This second agent should be capable of inhibiting biliary
 CC transport and may be a uridine 5'diphosphate glucuronyltransferase such
 CC as that encoded by the present sequence (designated UGRIA9). The method
 CC can be used in the treatment of cancer, gastrointestinal diseases and
 CC parasitic diseases.
 XX Sequence :320 BP; 608 A; 495 C; 547 G; 670 T; 0 other;
 SQ

Query Match 100.0%; Score 2320; DB 24; Length 2320;
 Best Local S.milarity 100.0%; Pred. No. 0;
 Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctcaagctgcagttctctgatggtggtgacaggggtgacccagccctctctctatgtgtg 60
 DB 1 ctcaagctgcagttctctgatggtggtgacaggggtgacccagccctctctctatgtgtg 60

QY 61 tctctgtctgctacgtgtggttgcagagggaggaagctactgttagtgcacatgat 120
 DB 61 tctctgtctgctacgtgtggttgcagagggaggaagctactgttagtgcacatgat 120

QY 121 gggagccactggttccactagatgctggtgtggaagaaactcattctcaggggggcatgag 180
 DB 121 gggagccactggttccactagatgctggtgtggaagaaactcattctcaggggggcatgag 180

QY 181 gtggttgcagtcacagagtgagttggcaactggaagatcactgaatgacacagt 240
 DB 181 gtggttgcagtcacagagtgagttggcaactggaagatcactgaatgacacagt 240

QY 241 aagacttattcaacttataccctggaagatgctggaccggggagttcaagggttttgc 300
 DB 241 aagacttattcaacttataccctggaagatgctggaccggggagttcaagggttttgc 300

QY 301 catgctcaatggaagcaagtcacagatatactattctctattaatgggttcacacat 360
 DB 301 catgctcaatggaagcaagtcacagatatactattctctattaatgggttcacacat 360

QY 361 gacattttgacttatttttttcaaatgacaggtttgtttaaagacaaaataatgta 420
 DB 361 gacattttgacttatttttttcaaatgacaggtttgtttaaagacaaaataatgta 420

QY 421 gaaactaaagagaggttcttgcagagttctctcgatctcttgcataactgtgac 480
 DB 421 gaaactaaagagaggttcttgcagagttctctcgatctcttgcataactgtgac 480

QY 481 ttaattgttgcacaaatatttctctccctccctccctccctccctccctccctccctcc 540
 DB 481 ttaattgttgcacaaatatttctctccctccctccctccctccctccctccctccctcc 540

QY 541 cactattctgaagaaggtgcacagtgcctgtgctctctctctctctctctctctctctct 600
 DB 541 cactattctgaagaaggtgcacagtgcctgtgctctctctctctctctctctctctctct 600

QY 601 tttaggttctcagatgcacatgactttcaagagagagtagtcaggaaccacatcatgcaacttg 660
 DB 601 tttaggttctcagatgcacatgactttcaagagagagtagtcaggaaccacatcatgcaacttg 660

QY 661 gagggaacttattatgacacagtttttcaaaaatgcctagaaatagcctctgaaatt 720
 DB 661 gagggaacttattatgacacagtttttcaaaaatgcctagaaatagcctctgaaatt 720

QY 721 ctccaaacacctgttacgagatgatctctacagccacacatcaaatcttggtgtgca 780
 DB 721 ctccaaacacctgttacgagatgatctctacagccacacatcaaatcttggtgtgca 780

QY 781 acgagcttctgttttgagatctcccaaacccgtgagcccaacatgatctcttctgtggt 840
 DB 781 acgagcttctgttttgagatctcccaaacccgtgagcccaacatgatctcttctgtggt 840

QY 841 atcaactgc-catacagggaaacgcgttgcctatggaattggaagcctacataatctctct 900
 DB 841 atcaactgc-catacagggaaacgcgttgcctatggaattggaagcctacataatctctct 900

QY 901 ggaacacatggaatgtgtttctctcttgggataatggtctcagaaattccagagaag 960
 DB 901 ggaacacatggaatgtgtttctctcttgggataatggtctcagaaattccagagaag 960

QY 961 aaagctatggaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
 DB 961 aaagctatggaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020

QY 1021 actggaacccagaccatcaatcttgcgaacacacagatacttgaagtggctaccccaa 1080
 DB 1021 actggaacccagaccatcaatcttgcgaacacacagatacttgaagtggctaccccaa 1080

QY 1081 aacgatctgttggttccacccgatgacccgtgacctttatccacccatgctgttcccatggt 1140
 DB 1081 aacgatctgttggttccacccgatgacccgtgacctttatccacccatgctgttcccatggt 1140

QY 1141 gtttatgaaagcatatgcaatggtgttcccatggtgagtgatgacccctgttctgtgacag 1200
 DB 1141 gtttatgaaagcatatgcaatggtgttcccatggtgagtgatgacccctgttctgtgacag 1200

QY 1201 atggacaaatgcacagcgcagtgagtaaggagctggagtgacccatggaatgttcttgaa 1260
 DB 1201 atggacaaatgcacagcgcagtgagtaaggagctggagtgacccatggaatgttcttgaa 1260

QY 1261 atgactctgaagatttagaataatgctctaaagcagtcacatgacaaaagtacaaag 1320
 DB 1261 atgactctgaagatttagaataatgctctaaagcagtcacatgacaaaagtacaaag 1320

QY 1321 gagaacatctgcgctctccagccttcaacaggaccgccggtggagccgtggacctg 1380
 DB 1321 gagaacatctgcgctctccagccttcaacaggaccgccggtggagccgtggacctg 1380

QY 1381 gccgtgtctgggtgaggtttgtgatgagcacaagggcgccacacctgggcccgca 1440
 DB 1381 gccgtgtctgggtgaggtttgtgatgagcacaagggcgccacacctgggcccgca 1440

QY 1441 gccacacacacacacacacacacacacacacacacacacacacacacacacacacac 1500
 DB 1441 gccacacacacacacacacacacacacacacacacacacacacacacacacacacac 1500

QY 1501 gtcgtgtgcagagtgaccttcatcaccttttaaatgttgccttatggctaccggaaaatgc 1560
 DB 1501 gtcgtgtgcagagtgaccttcatcaccttttaaatgttgccttatggctaccggaaaatgc 1560

QY 1561 ttgggggaaaaagggcgaggttaagaaagcccaacaaatcccaagacccattgagaagtgggt 1620
 DB 1561 ttgggggaaaaagggcgaggttaagaaagcccaacaaatcccaagacccattgagaagtgggt 1620

QY 1621 gggaaaataaggttaaaattttgaacattcccttagtcaatttcccaacttgaacacagaatc 1680
 DB 1621 gggaaaataaggttaaaattttgaacattcccttagtcaatttcccaacttgaacacagaatc 1680

QY 1681 agtgttaaaatcaatttttttatttaaggaataactttgataaaataatgaagcccaag 1740
 DB 1681 agtgttaaaatcaatttttttatttaaggaataactttgataaaataatgaagcccaag 1740

QY 1741 agtgtttaaaaaatctcttaataaaaaataatagactcgtcgtcgtcgtcgtcgtcgtcgt 1800
 DB 1741 agtgtttaaaaaatctcttaataaaaaataatagactcgtcgtcgtcgtcgtcgtcgtcgt 1800

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 06:34:42 ; Search time 682.45 Seconds
(without alignments)
5836.679 Million cell updates/sec

Title: US-09-835-082-1
Perfect score: 2320

Sequence: 1 ctacagtcagttctctgat.....ttctgatgttctctacaact 2320

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2320	100.0	2320	24	ABA05198		Human uridine 5'di
2	2320	100.0	2320	24	ABA05199		Human uridine 5'di
3	1581.4	68.2	2351	13	AAQ27369		HUG-Brl. Homo sap
4	1576.8	68.0	2368	13	AAQ27370		HUG-Brl. Homo sap
5	1551.8	66.9	1655	23	AAS91252		DNA encoding novel
6	1004.4	43.3	18887	24	AAS18543		DNA encoding UDP g
7	998.6	43.0	1001	21	AAH51403		Human UGT1A7 relat
8	870.6	37.5	1001	21	AAH51404		Human UGT1A7 relat
9	798	34.4	1667	13	AAQ33027		Exon 5 from the US

10	769.6	33.2	1001	21	AAH51405	Human UGT1A7 relat
11	766.6	33.0	930	21	AAZ45116	UDP-glucuronosyltr
12	741.2	31.9	951	21	AAZ45115	UDP-glucuronosyltr
13	735	31.7	735	21	AAZ45118	UDP-glucuronosyltr
14	659.8	28.4	759	21	AAZ45117	UDP-glucuronosyltr
15	578.8	24.9	874	24	ABL01468	Murine apoptosis i
16	364.2	15.7	391	20	AAZ07208	Human lung tumour
17	364.2	15.7	391	21	AAC79066	Human lung tumour
18	364.2	15.7	391	23	AAD23141	Human lung tumour
19	359	15.5	1001	21	AAH51406	Human UGT1A7 relat
20	353	15.2	1001	21	AAH51407	Human UGT1A7 relat
21	345.2	14.9	1854	21	AAZ95200	Human UDP-glucuron
22	345	14.9	1766	23	AAS91380	DNA encoding novel
23	344.4	14.8	2107	19	AAV15900	Uridine diphospho-
24	342	14.7	380	21	AAC98723	Human colon cancer
25	339.8	14.6	2092	21	AAZ95199	Human UDP-glucuron
26	333.2	14.4	1976	21	AAZ95206	Human UDP-glucuron
27	332.8	14.3	1650	21	AAC65396	Human carbohydrate
28	327.2	14.1	2802	23	AAS69711	DNA encoding novel
29	325.6	14.0	2111	23	AAS69712	DNA encoding novel
30	322.4	13.9	1636	24	AAD24666	Human drug metabol
31	322.4	13.9	1722	22	AAD06860	Human breast cancer
32	322.4	13.9	2974	22	AAS46185	Human DNA encoding
33	318	13.7	1859	23	AAS69710	DNA encoding novel
34	314.6	13.6	1713	22	AAD06844	Human breast cancer
35	302.4	13.0	508	21	AAC98478	Human colon cancer
36	296	12.8	1001	21	AAH51408	Human UGT1A7 relat
37	282	12.2	2025	23	AAS91253	DNA encoding novel
38	230.4	9.9	366	21	AAH87085	Rat hepatocyte car
39	222.4	9.6	775	13	AAQ33026	Exons 3-4 from the
40	220.8	9.5	273	21	AAH87491	Rat hepatocyte car
C 41	203	8.8	2448	23	AAZ72022	DNA encoding novel
C 42	203	8.8	2448	23	AAS74393	DNA encoding novel
C 43	203	8.8	2448	23	AAS78611	DNA encoding novel
C 44	203	8.8	2448	23	AAS78773	DNA encoding novel
C 45	203	8.8	2448	23	AAS84057	DNA encoding novel

ALIGNMENTS

RESULT 1
ABA05198
ID ABA05198 standard; cDNA; 2320 BP.
XX ABA05198;
AC ABA05198;
XX 04-MAR-2002 (first entry)
DE Human uridine 5'diphosphate glucuronyltransferase UGT1A9 cDNA #1.
XX Human; uridine 5'diphosphate glucuronyltransferase enzyme; UGT1A9;
KW flavopiridol; cancer; gastrointestinal disease; parasitic infection;
KW cytosolic; ss.
XX Homo sapiens.
OS Homo sapiens.
XX Key Location/Qualifiers
FT CDS 19..1611
FT FT /*tag= a
FT FT /product= "UGT1A9"
XX WO200180896-A2.
XX 01-NOV-2001.
XX 12-APR-2001; 2001WO-US12526.
XX 21-APR-2000; 2000US-0553829.
XX (ARCH-) ARCH DEV CORP.
XX Retain MJ, Innocenti F, Iyer L;

APPLICANT: BELKNAP, WILLIAM R
TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE
TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: NANCY J. PARSONS
STREET: 800 BUCHANAN ST.
CITY: ALBANY
STATE: CA
COUNTRY: USA
ZIP: 94710
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/797,226
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PARSONS, NANCY J
REGISTRATION NUMBER: 40,364
REFERENCE/DOCKET NUMBER: 0011.97
TELEPHONE: (510) 559-5731
TELEFAX: (510) 559-5777
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1607 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Solanum tuberosum
STRAIN: cv. Lemhi Russet
INDIVIDUAL ISOLATE: SGT1750
DEVELOPMENTAL STAGE: mature, somatic
TISSUE TYPE: tuber
IMMEDIATE SOURCE:
LIBRARY: lambda gtl1 cDNA library
CLONE: SGT 1750
FEATURE:
NAME/KEY: CDS
LOCATION: 20..1486
OTHER INFORMATION: /product= "solanidine
PUBLICATION INFORMATION: glucosyltransferase"
AUTHORS: Moehs, Charles P.
AUTHORS: Allen, Paul V.
AUTHORS: Friedman, Mendel
AUTHORS: Belknap, William R.
TITLE: Cloning and expression of solanidine
TITLE: UDP-glucose glucosyltransferase from potato
JOURNAL: The Plant Journal
VOLUME: 11
ISSUE: 2
DATE: 1997
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 488
US-08-797-226-1

Query Match 1.9%; Score 44.8; DB 2; Length 1607;
Best Local Similarity 53.4%; Pred. No. 0.0072;
Matches 94; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
Qy 1068 gtggctaccccaaacgactgctggtccacccgagccggtccctttatccaccatgc 1127
|||||
Db 1069 GTGGGTCCACACGCTTACCATCGGACATTCAGCAACAGCGGGTTTCATGCTATG 1128
|||||

Qy 1128 tgggtcccatggtgtttatgaagcatatgcattgcttcccatggtgatgcctt 1187
|||||
Db 1129 TGGTACTAATTCGGTTCGGAAGCCATCACTTTTGGCGTGCATGATACATGCCACT 1188
|||||
Qy 1188 gtttggtgatcagatggacaaagcgcatggagactaaaggagctggagtgga 1243
|||||
Db 1189 TTATGCTGATCAATTCTACAACGAGAAGGTAGTCGAGGTTAGGGATTGGGAATCA 1244
|||||

RESULT 15
PCT-US92-00282-37
; Sequence 37, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US92-00282-37

Query Match 1.9%; Score 43.6; DB 5; Length 51;
Best Local Similarity 92.0%; Pred. No. 0.0021;
Matches 46; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1444 cagcagctcacttggtaccaggtaccattccttggtgacgtgattggttctt 1493
|||||
Db 1 CAGGACCTCACCTGGTTCAGTACCACCTCTTTGGATGCTGATTGTTCTCT 50
|||||

Search completed: August 21, 2002, 11:59:18
Job time: 32806 sec


```

; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
PCT-US92-00282-18

```

```

Query Match          7.0%; Score 161.8; DB 5; Length 1190;
Best Local Similarity 51.8%; Pred. No. 1.3e-36;
Matches 414; Conservative 0; Mismatches 382; Indels 3; Gaps 2;

```

```

QY 68 tgcctgacccgtgcttccgagcagcaggaagcactactagtgcccatgagatggagacc 127
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140 TGCTGGGCCAGTGTGTCCCATGCTGGGAGATGACTGTGTGATCCAGTGGATGGCAGCC 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 actggttcaaccatgagtgctggtggtgagaaactcattcagggggccatgaggtggtg 187
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 ACTGGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGAGAGGGGACATGAATAGTTG 259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 188 tagtcagccagagtgagtggtggcaactgggaagatcacatgaaatgacagtgaaagact 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 260 TCCTAGCACCTGACGCTCGTGTGTACATCAGAGCGGAGCATTTTACACCTTGAAGAGCT 319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 attcaactcattatcaccctggaggtatctggaccggaggt--caaggctttggccatgc 305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 320 ACCCTGTGCCAATCCAAAGGAGGAGTGTGAAGAGTCTTTTGTGTAGTCTCGGGCATATG 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 306 tcaatggaagcacagtagtaagatatattctcttataatggttccatcaaatgacat 365
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 TTTTGTAGATGATCTTCTCGCAGCGTGTGATCAAAACATACAAAGAAATAAAGG 439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 366 tttgact-tattttttcaattgagaggtttgtttaagacacaaataattagtagaat 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 440 ACTGTGATGCTTTTGTGTGCTGTGCCACTTACTGCAACAAGAGGAGCTCATGGCCT 499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 425 acttaaggagaggtttctttgatgcaggtttctctgatccctttgtataactgtgcttaa 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 500 CCTGCGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCTTTCTTCTTCAGCCCCA 559
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 485 ttgttgccaaattctccctccctccctggtgtcttcgcaagggaataattggcaact 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 560 TGTGGCCACGACTGCTCTGCCACACTGTATCTTCTGATGACACTGCCATGCAAGCC 619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 545 atcttgaaagaggtgcagtgccctgctcctcttctctctatgtcccccagaattctcttag 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 620 TGGAAATTGAGGCTACCAAGTGCCTCCAAACCCATTTCTCTACGTGCCAGGCCCTCTCTCT 679
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 605 gttctcagatgccatgactttcaaggagagagtagcaggaacacatcatgcaactggaggg 664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 680 CTCATTCAGATCAGATGACCTTCTGCGGGGTGAAGACATGCTCATTTGCTTTTTCAC 739
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 665 aacattattatgccaccggtttttcaaaaatgcccctagaaatagcctctgaaattctcc 724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 740 AGAATTTCGTGCGACGTGTTTATTTCCCGTATGCAACCTTGCTTCAGAAATTCCTTC 799
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 725 aaacacctgttaagagtagatgatctctacagcacaacataatttgggttgggaacagg 784
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 800 AGAGAGAGGTGACTGTCCAGACCTATGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 859
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 785 actttgtttggactatcccaaacccgtgtatgcccacacatgatctctctggtgatca 844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 860 ACTTTGTGAAGGATGACCTAGGCCCATCATGCCCAATATGTTGTTTGTGGTGAATCA 919
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 845 actgcatcaggaagacc 863
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 920 ACTGCCCTTCCAAAATCC 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 10
PCT-US92-00282-12
; Sequence 12, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.

```

```

; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARRY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1197 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
PCT-US92-00282-12

```

```

Query Match          6.9%; Score 161.2; DB 5; Length 1197;
Best Local Similarity 52.6%; Pred. No. 2e-36;
Matches 421; Conservative 0; Mismatches 373; Indels 6; Gaps 3;

```

QY 85 gccagagcaggaagcactgtagtgcccatggatggagggccactgg--ttcaccatga 142
Db 304 GCTGAGAGTGGAAAGGTGTTGGTGGTCCCATGATGSCAGCCCTGTCTCAGTCATGC 363

QY 143 ggtcgtgtgtggaacacatctcaggggcatgaggtgtttagtcacatgcacagg 202
Db 364 GGGAGGCTTGGGGAGCTTCATGCGCAGAGGCCACAGCGGTGGTCTCACCACAGG 423

QY 203 tgagttggcaactgggaagatcactgaattgcacagtgaaagacttattcaacttcata 262
Db 424 TGAATATGCACATCAAGAAGAGAAATTTTTCACCTGACAGGCTATGCTGTTCATGGA 483

QY 263 ccttgagagatctgcacggaggttcaaggttttggccatgctcaat---ggaagcac 319
Db 484 CCAGAGAGAAATTTGATCGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543

QY 320 aagtaagagatataattctctatattatgggttccatcaacatgacatttttgacttattt 379
Db 544 ATCTTCTGAAGAGATATTCTAGAAGTATGGCAATTTATGAACAATGTATCTTTGGCCCTC 603

QY 380 tttaaatgicagaggtttgttttaaagacacaaataattagtaacttaagaggaggtt 439
Db 604 ATAGGTGTTGTTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGATGCTACTT 663

QY 440 cttttagatcaggtttctctctctctctctctctctctctctctctctctctctctct 499
Db 664 CCITTGATGTTTAAACAGACCCCGTTAACCTCTGTGTGGGGGGGTGCTGGCTAAGTACC 723

QY 500 tctccctccctccctggtctctctctctctctctctctctctctctctctctctctct 559
Db 724 TGTCGATTCTCTCTGTGTTTTTTTGGAGGTATCTCCATGCTAGACTTTAGAGCCCA 783


```

QY 1505 tctgacagtgccctcatcaccctttaaagtgtgtgtcttattggtaccgggaaatgcttgg 1564
D 1496 TGCTGACAGTGGCGCTTCATACCTTTAAATGTTGTCTATGGCTACCGGAAATGCTTGG 1555
QY 1565 ggaataaaggcgagtgaaagaaagccacaaatccaaagaccattgaaagtgggtggga 1624
D 1556 GGAARAAAGGCGGAGTTAGAAAGCCACAAATCCCAAGNCCNTGAGAAAGTGGGTGGGA 1615
QY 1625 aataaggtataaatttgaaccattccctagctatttccaaacttgaacagaaatcagtg 1684
D 1616 AATAAGGTAAATTTTGAACCATTCCTTAGTCTATTTCCAACTTGAACAGAAATCAGTG 1675
QY 1685 ttaattcatttattcttattagaaatcatttgcataataatcaagcccccagagtg 1744
D 1676 TTAATATCATTTTATTATTAGGAATACATTTTGCATAAATTAATACGCCCCAGAGTG 1735
QY 1745 ctttaaaaaattctcttaataaaaaataatagactcgtctagtcagtaagaatatttgat 1804
D 1736 CTTTAAAAAATCTCTTAATAAATAAATAATAGACTCGCTAGTCAATGAAGATATTGAAT 1795
QY 1805 atgtatcgtcccccctccggtgtcttcttgatcaggtgacatgtccatttttcagagac 1864
D 1796 ATGTATCTGCGCCCTCGGCTGCTTTGTATCAGGATGACATGTGCCATTTTTCAGAGAC 1855
QY 1865 gtgcagacaggtcggcattctagattactttcttacttgaacatggcctgtttggga 1924
D 1856 GTGCAGACAGGCTGGCATTTCTAGATTACTTTTCTTACTCTGAAACATGGCGCTGTTGGGA 1915
QY 1925 gtgcgggattcaagggtgcccaccgctgcccctactgcaaaatggcagttttaaactta 1984
D 1916 GTGCGGATTCAAAGGTGTCCACCGCTGCCCTTACTGCAATGGCAGTTTAACTTTA 1975
QY 1985 tcttttggcttcgagatggttgcaattgatcctttaaaccataatggcagtcctcatc 2044
D 1976 TCTTTGGCTTCTGACAGATGGTGGCAATGATCTTTAACCAATATGGTCAGTCTCATC 2035
QY 2045 tctgtccttcttcattagggtgcacccttgtgttttaagaagggagagcttttaccctta 2104
D 2036 TCTGTCTGCTTCTATAGTGTGCCACCTGTGTGTTTAAAGAGGGAAAGCTTTGTACCTTA 2095
QY 2105 gagttagtgaaatgaatgaatgagcttgagtgacactgaacagacatatgattcttg 2164
D 2096 GAGTGTAGTGAATGAATGAGCTTGAGTGTGACTGAGAACAGCATATGATTTCTTG 2155
QY 2165 ctttgggaaagaatgatgctatgaattggtggtggtgtatttggagaagataatca 2224
D 2156 CTTTGGGAAAAAGAAATGATGCTATGAAATTTGGTGGTGGTATTTGAGAAGATAATCA 2215
QY 2225 ttgcttatgtcaatgagctgaatttgataaaaaaccacaaatacagctatgaagtctg 2284
D 2216 TTGCTTATGTCAATGGAGCTGAATTTGATAAAAACCCCAAAATACAGCTATGAGTGTG 2275
QY 2285 ggaagtttactttttcttgatgttcttcacaact 2320
D 2276 GGCAAGTTTACTTTTTTCTGATGTGTTCTCTACAACT 2311

```

RESULT 3

```

PCT-US92-00282-25
; Sequence 25, Application PC/US9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.

```

```

; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1561 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US92-00282-25

```

Query Match 41.7%; Score 967.2; DB 5; Length 1561;

Best Local Similarity 98.0%; Pred. No. 3.3e-268;

Matches 1011; Conservative 0; Mismatches 18; Indels 3; Gaps 3;

```

QY 1291 aaagcagtcacatcaagacaaagttacaagagaaacataatgcgcctctccacccctaac 1350
D 248 AACGAAACTGTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 307
QY 1351 aagagaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1410
D 308 AAGGACCGCCGCGTGGAGCGCTGGACCTGGCGCTGTCTGGGTGGAGTTGTGTGTGTGTGT 367
QY 1411 cacaagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1470
D 368 CACAAGGGCGCGCCACACCTGGCGCCCGCAGCCACGACCTACCTGGTACCAAGTACCAT 427
QY 1471 tcccttgacgtgattggtttccctcttggcgcgtgctgct-gacagtgcccttcacacct 1529
D 428 TCCTTGGACGTGATGTTGTTCTTCTTGGCGCTGTGGCTGGACAGTGGCTTTCACCTT 487
QY 1530 taaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1589
D 488 TAAATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 547
QY 1590 ccacaaatccaaagaccacatgagagtggtgggaaataaggttaaaatttggaccattc 1649
D 548 CCACAAATCCAAAGACCCATTTGAGAAGTGGGTGGGAAATTAAGCTAAATTTTGAACCATTC 607
QY 1650 cctagtcatttccaaacttgaacacagaaatcagtgtaaatctatttattcttataag 1709
D 608 CCTAGTCATTTCCAACTTGAACACAGAAATCAGTGTAAATTCATTTTATTTCTATTAAAG 667
QY 1710 gaaatacttctgataaataatcagccccagagtgctttaaaaaattctctctataaata 1769
D 668 GAAATFACTTTGCATAAATTAATCAAGCCCGCCAGAGTGGCTTTAAAAAATTTCTTTAAATAAA 727
QY 1770 aataatagctcgtactcagtaaaagatatttgaatgtatcgltg-ccccctccggtgct 1828
D 728 ATAATAGACTCGGTAGTACGTAAAGATATTTGAATATGTATGTGTGCCCCCTCTGGGTCT 787
QY 1829 ttgatcagatgacatgtgcccatttttcagagcggtgcagacaggtggtgacattcaga 1888
D 788 TTTGATCAGGATGACATGTGCCATTTTTCAGAGAGCGTGCAGACAGAGCTGG-ATTCTAGA 846
QY 1889 ttacttttctactcgaacatgacctgttttgggagtgcggtgattcaagggtggtccca 1948

```


Qy 141 gaggtcggtggagaaactcattctcaagggggagtgagggtgtgttagtcatgccaag 200
Dy 132 GGGGAGGCGCTGGGGAGCTCCATGCGCAGAGGCCACCGAGGGGTGGTCTCACCACCA 191
Qy 201 ggtgagttggcaactgggaagtaactcaatgacacagtgagacttatcaactcata 260
Dy 192 GGTGAATATGCATCAAGAGAGAAATTTTCACCTGACAGCCTATGCTGTCCATG 251
Qy 261 taccctggagatctggacggaggttcaagggttttggccatgctcaat---ggaagc 317
Dy 252 GACCAGAGGGAATTTGATCGGGTTACGGCTGGCTTACACTCAAGGGTCTTTGAAACAGA 311
Qy 318 acaagtcagagatatactctcttaataatgggttcatacaatgacattttgacttatt 377
Dy 312 ACATCTTGAAGAGATATCTAGAGATATGGCAATTAAGCAATGATCTTTGGCCCT 371
Qy 378 ttttcaaatcagagaggtttgtttaagagaaaaaattagtagaataacttaagagag 437
Dy 372 TCATAGGTGTTGTGGAGCTACTGCATAATAGGCCCTGATCAGCAACCTGAAATGCTAC 431
Qy 438 tttctttgatgcagtttttctcgatccttttgataactgtgcttaattgttgcacaata 497
Dy 432 TTCTTTGATGTTTAAACAGACCCGTTAACTCTCGGGGGGTGCTGCTTAAGTA 491
Qy 498 tttctccctccctccggtgtgttcgtccaggggaatactttggcaactatcttgaagagg 557
Dy 492 CTGTGCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCAGTGTGACTTAGACTTTAAGG 551
Qy 558 tgacagtgccctgtcctctctctctatgtcccccagaaactctcttagggttccagatgc 617
Dy 552 CACAAGTGTCCAAATCTCTCTCTATATCTTAATCTTAAGTTACTTAAGCACTTTCAGACA 611
Qy 618 catgacttcaagagagagagtagcaggaaccacatcatgacttgagagaaacttattatg 677
Dy 612 CATGACATCTCGCAAGGGTCAAGAACATGCTCTACCTCTCGGCCCTGCTTACATATG 671
Qy 678 ccacggtttttcaaaaatgccttagaaaaagcctctgaaatctcccaaacactgttac 737
Dy 672 CCATACTTTTTCTGGCCCTTATGCAAGTCTGCTCTGAGCTTTTTCAGAGAGAGGTGTC 731
Qy 738 ggaatgactctatcagcacacacatcaatttggttggtagaagcaacttatttggga 797
Dy 732 AGTGTGATCTTGTGAGTATGCAATGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTG 791
Qy 798 ctatcccaaacccgtgatcccaacatgatctctcattggtggtatcaactgccaatgag 857
Dy 792 CTACCCAGGCGGATCATGCCAACATGGTCTTCAATTTGGGGCATCAACTGTGCCAACGG 851
Qy 858 aaagccgttgccatggaatttgaagcctacattaatgctctctgagaaatggaattgt 917
Dy 852 GAAGCCACTATCTCAGGAATTTGAAGCCTACATTAATGCTTCTGAGAACATGGAATGT 911
Qy 918 ggtttctctttgggataatggttcagaaattccagagaaagcaactatggaattgc 977
Dy 912 GGTTTCTCTTTGGGATCAATGGTCTCAGAAATTCAGAGAAAGCAATGCAATTCG 971
Qy 978 tgatctttggcaaaatccctcagacagtcctctggtggtgagctacactgaaacccacatc 1037
Dy 972 TGATCTTTGGGCAAAATCCCTCAGACAGTCTCTGCTGGCGGTACACTGGAACCCGACATC 1031
Qy 1038 gaattctgcaacaacacgatacttggtaagtgggttaccaccaaacgactctgttggtca 1097
Dy 1032 GAATCTTGGCAACAACACGATACTTTGTTAAGTGGCTTACCCCAAAACGATCTGTGCTCA 1091
Qy 1098 cccgatgacccgtgcctttatcaccatgctgttcccatggtgtttatgaagagcatag 1157
Dy 1092 CCGATGACCGTGCCTTTATCACCCTATGCTGTTCCATGGTGTATGATGAAGCATATG 1151
Qy 1158 caatggcgttcccatggtgatgacctgtttgttgggtgatcagatggacaatcaagag 1217
Dy 1152 CAATGGCGTTCATGATGCCCTGTTTGGTGTGATCAGATGGCAATGCAAGCG 1211

Qy 1218 catgagactaaggagctgagtgacccctgaatgttcttgaaatgacttctgaatatt 1277
Dy 1212 CATGAGACTAAGGGAGCTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1271
Qy 1278 aaaaaatgctcttaaaagcagtcataatgacaaagtacaaagagagacatcatgcgct 1337
Dy 1272 AGAAATGCTCTTAAGAGCAGTCAATGACAAAGCTTACAAAGAGAACATCAGCGCCT 1331
Qy 1338 ctccagccttcaaaagacccgcccgggtggagccgctggagccttggccgtgttcttgggtgaa 1397
Dy 1332 CTCCAGCCTTCAAAAGAGCCGCCGGTGGAGCCGCTGGACCTGGCCCGTGTCTTGGGTGA 1391
Qy 1398 gtttgtgatgagcacaagggcgqccacacactgcgcccgcagccacacacactcactcg 1457
Dy 1392 GTTGTGTGAGGAGCACAAGGGCGGCCACACCTGCGCCCGCAGCCACCACTCACCTG 1451
Qy 1458 gtaccagtcacattccttgagcgtgattgttctcctctggcgtgctgtgacagtggc 1517
Dy 1452 GTACCAGTACCATTCTCTTGGAGCTGATTTGGTTTCTTGGCCGCTGCTGTGACAGTGGC 1511
Qy 1518 ctcaatcacctttaaatgttgtcttattggttaccggaatgcttgggaaaaaagggcg 1577
Dy 1512 CTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGAAAAAAGGGCG 1571
Qy 1578 agttaagaagcccaacaataccaagaccatttgagaagtgggtgggaaaaaaaggttaaaat 1637
Dy 1572 AGTTAAGAAAGCCCAAAATCCAGACCCATTTGAGAAGTGGGTGGGAAATAAGTAAAAAT 1631
Qy 1638 ttggaacattcctctagtcatttccaaacttgaaacagaaatcagtggttaaaattcatttt 1697
Dy 1632 TTTGAACCATCTCCCTAGTCAATTTCCAACTTGAACAGAACTAGTAAATTCATTTT 1691
Qy 1698 attcttattagaataactcttgataaaataatcaagcccaagagtgctttaaataatcc 1757
Dy 1692 ATCTCTATTAGGAAATAGCTTTGCAATAATTAATCAGCCCCAGAGTGTCTTAAAAAATTC 1751
Qy 1758 tcttaataaaaaataactcgtcagtcagtaaaagataattgaaatgattatcgtgccc 1817
Dy 1752 TCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATATGATATCGTCCC 1811
Qy 1818 cctccggtgcttgatgaagtagacatgtgccattttcagagagcagtcagacagct 1877
Dy 1812 CTTCCGGGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGAGCTGCACACAGCT 1871
Qy 1878 ggcattcttagattactttcttactctgaacatggcctgttgggagtgccggattcaa 1937
Dy 1872 GGCATTTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTGGGAGTGGGGATTCAA 1931
Qy 1938 agtgtgtccacgcgtccctcactgcaaatgacagttttaaacttcttcttcttctct 1997
Dy 1932 AGGTGTCCTCCCGCTGCCCCCTACTGCAATGGCAGTTTTTAATCTTATCTTTTGGCTTCT 1991
Qy 1998 gcagatggttgcaattgactcttaaccaataatggtcagtcctcactctctgctcttc 2057
Dy 1992 GCAGATGTTGCAATTTGATCTCTTAACCAATAATGGTCAGTCTCATCTCTGCTCTCTTC 2051
Qy 2058 atagtgccaccttggtgtttaaagaagagcctttagccttttagagtgtagtgaa 2117
Dy 2052 ATAGTGTCCACCTTGTGTTTAAAGAGGAGGAACTTTTACTCTTATAGAGTGTAGTGAA 2111
Qy 2118 atgaatgaatggttgagtgacactgagaacacatgatatttcttcttggggaaaaa 2177
Dy 2112 ATGAATGAATGGTGTGAGTGCACCTGAGAAACAGCATATGATTTCTTGGGGAAAAA 2171
Qy 2178 gaatgactatgaatgtgtgggtggttatitgagaagataaatacttcttctatcaa 2237
Dy 2172 GAATCATGCTATGAATTTGGTGGGTGTGATTTTGAAGAGATATATCATTCATTCATA 2231
Qy 2238 atggagctgaatttgataaaaaaaccccaaaaatacagctatgaagtgctgggcaatttactt 2297
Dy 2232 ATGGAGCTGAATTTGATAAAAAACCCAAAAATACAGCTATGAAGTGTGGCAAGTTTACTT 2291
Qy 2298 ttttctgtgttctcctaact 2320

ORIGIN

Query Match 27.8%; Score 644; DB 10; Length 723;
 Best Local Similarity 96.7%; Pred. No. 1.1e-157;
 Matches 700; Conservative 0; Mismatches 20; Indels 4; Gaps 4;

QY 1078 caaacagatctgttgcacccgatgaccogtgcctttatcacccatgctggttcccat 1137
 Db 2 CAACAGATCTCTGTGGTCAACCGGATGACCGTG-CTTTATCACCATGCTGTGTCAT 60
 QY 1138 ggtgtttatgaagcatatcaatgctggttcccatggtgatgcttctgttggatg 1197
 Db 61 GGTGTTTATGAAGCATATCAATGCGGTTCCTCCATGGTGATGCCCTGTGTTGGTAT 120
 QY 1198 cagatggacaaatgcaagcgcgatgagacataagggagctggagctgaatgtctg 1257
 Db 121 CAGATGGACAAATGCAAGCGCATGAGAGCTAAGGGAGCTGGAGTACCCCTGAATGTTCTG 180
 QY 1258 gaaatgactctgaagatttagaaaatgctctaaaagcagtcacatcaatgacaaaagtac 1317
 Db 181 GAATGACTTCTGAAGATTTAGAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTAC 240
 QY 1318 aagagagacatcatcgctctccagccttcaagagacgcgcgctggagccgctggac 1377
 Db 241 AAGAGAACATCATCGCCTCTCCAGCCTTCACAGAGACGCCCGGTGGAGCCGCTGGAC 300
 QY 1378 ctggcgcgtgtctgtggtgagttgtgtagagcacaagggcgccacacacctgccc 1437
 Db 301 CTGGCCGTGTCTGGTGGAGTTGTGATGAGGCACAGGCGGCCACACCTGCGCCCC 360
 QY 1438 gcagccacagacctcaactggtagcagtagcaccattccttgagcgtgattggttcccttg 1497
 Db 361 GCAGCCACAGACCTCACCTGTTACCAATCCATTCCTGGACGTGATGGTTCTCTCTTG 420
 QY 1498 gccctgctgacagtgctctcacccttcaactttaaattgtgtctatgctaccgaaa 1557
 Db 421 GCCCTGTCTGACAGTGGCCTTCATCACCTTTAAATGTTGCTTATGGCTACCGGAAA 480
 QY 1558 tgcctgggaaaaagcagcttaagaaagcccaaaatccaaagaccattgagaagt 1617
 Db 481 TGCTGGGGAAGAAAGGCGAGTAAAGAACCCCAAAATCAAGACCCCTTGAGAGTG 540
 QY 1618 ggtgggaaataaggtataaa-ttttgaaaccattcccttagtcaatttccaaacttgaaaaacag 1676
 Db 541 GGTGGGAAATAAGGTAAATTTTGAACCAATTCCTAGTCAATTCCTCAAACTTGAACACAG 600
 QY 1677 aatcagtgtaaatcattttattcttatttaaggaatacttgcataaataatcagcc 1736
 Db 601 AATCAGTGTACATTCATTTTATTTATTAAGGAATAC-TTGCATATAATTAATCAG-C 658
 QY 1737 ccagagtgtttaaaaaattctcttaaaataaaataatagactcgtagtcaagaaat 1796
 Db 659 CCAGAGTGTCTTAAAAAATTTCTTAATAAATAATAGACTCGTAGTCAAGATTTGAT 718
 QY 1797 attt 1800
 Db 719 ATGT 722

RESULT 14
 BG469506
 LOCUS 602532818F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4660540 5',
 DEFINITION mRNA sequence.
 ACCESSION BG469506
 VERSION BG469506.1 GI:13401781
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 829)

AUTHORS
 TITLE
 JOURNAL
 COMMENT

NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: L1CM1458 row: c column: 05
 High quality sequence stop: 795.

FEATURES

source

1..829

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4660540"

/clone_lib="NIH_MGC_15"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pOTB7; Site: 1; XhoI; Site: 2;

EcoRI; CDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adapter: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 186 a 190 c 207 g 246 t

ORIGIN

Query Match 27.8%; Score 641.2; DB 10; Length 829;
 Best Local Similarity 89.9%; Pred. No. 6.4e-157;
 Matches 732; Conservative 0; Mismatches 78; Indels 4; Gaps 4;

QY 5 gctgcagttctctgattgctgacaggtgacagcccttctctctatgtgtgtc 64
 Db 18 GCTGCAGTTCTCTCATGCTCGCGCAGGGTGCACGCCCGTCTTATGTGTGTGTC 77
 QY 65 tgcgtgcacctgtgcttgcagagggaggaagctactgtagtgcacatggatgga 124
 Db 78 TACTGTGACCTGTGGCTTGGCGAGGAGGAAAGCTGTGGTAGTGCCCATGTATGGA 137
 QY 125 gccactggttccacctgaggtcggtggtggaagaaactctcagggggcaggtgg 184
 Db 138 GTCAGTGGTTACCATGTCAGTGTGGTGGAGAAACTTATCTCAGGGGCGATGAGTGG 197
 QY 185 ttgtagtcaatgcagaggtgagttggcaactgggaagatcaactgaattgcacagtgaaga 244
 Db 198 TTGTAGTCAATGCGCAGAGTGAGTTGCAACTGGAAAGATGACATGAATTCACAGTGAAGA 257
 QY 245 ctattcaacttcatataccctggaggtctggaacggggagttcaaggcttttgcacctg 304
 Db 258 CTTACTCAACCTCGTACACTCTGGAGATCAGAACCGGGAATTCATGTTTTCGCCCATG 317
 QY 305 ctoaatggaagacacaaatgacgaagtatatctcttataatgggttccatacaatgaca 364
 Db 318 CTCATGGAAGACACAGGCACAAAGATATATTTTCTCTATTAATGAGTTTCATCCAGTGGTT 377
 QY 365 ttttgcactatttttttcaaatgcagaggttcttttaagagacaaaaaattagtagaat 424
 Db 378 TTCTTGACTTATTTTTCGCATTCGAGGAGTTGGTTTAAATGACCCGAAAATTAGTAGAAT 437
 QY 425 acttaagagagaggttcttttgatgcaggttttc-tcgatccttttgaataactgtggtta 483
 Db 438 ACTTAAAGGAGAGTCTCTTTGGATGACAGTGTTCGTGGATCCTTTTGTATACCTGTGGCTTA 497
 QY 484 attgttgcaaaatatttctccctccctccctcggtgtcttcgcagggggaataactttggcac 543
 Db 498 ATTGTGCTGAATATTTCTCCCTCCCTCTGTGGTGTCTTCACCGGGGAATATTATTGGCCAC 557


```

Db 181 GGAGACTAAGGAGCTGGAGTGCACCTGAACTTCTGGAATGACTTCTGAAGATTAGA 240
QY 1281 aaatgctctaaagcaatcatcaatacaaaaattacaagaaacacatcatgcctctc 1340
Db 241 AAATGCTCTAATGAGCAGTCAATGACAAAAGTTTACAGGAAGACATCATCGGCTCTC 300
QY 1341 cagccttcaacaggaccgcccgggtgagccgctggacacctggccggttcttgggtggagt 1400
Db 301 CAGCCTTCAACAGGACCGCCGGTGGAGCGCTGGACCTGGCCGCTTCTGGGTGGAGTT 360
QY 1401 tctatgagcacaagccgcccacacacctgcgcccgcagccacacacctcaactagta 1460
Db 361 TGTATGAGGACACAGGCGCGCCACACCTGCGCCCGCCAGCCACGACTCACCTGGTA 420
QY 1461 ccagtaacctcttggcagtgatgtgttctcttggcgttggcgttggcgttggcgtt 1520
Db 421 CCAGTACCATCTTGGACGTGATGTTGCTTCTTGGCGCTGCTGACAGTGGCCTT 480
QY 1521 catcactttaaatgttgtcttatgctaccgaaatgcttgggaaaaaaggccagtt 1580
Db 481 CATCACCCTTAATGTTGTGCTTATGCTACCGGAATGCTTGGGAAAAAAGGCGAGT 540
QY 1581 taagaagcccaacaaatccaagaccattgagaagtggtgggaaataaggtaaatctt 1640
Db 541 TAAGAAAGAGCCCAACAAAGACCCATTGAGAAGTGGGTGGGAAATAAGGCTAAATTTG 600
QY 1641 gaaccattccct-agtcatcttccaaacttgaacacagaaatcagtgtaaattcatctt 1699
Db 601 GAACCATCTCCCTAAGTCAATTCACAACTTGAAACAGAAATCAAGTGTCAATTCATCT -A 658
QY 1700 tcttattaagaaatacttgcataataataacagcccgagtgctttaaataattctc 1759
Db 659 TCTTATTACGAAATACTTCGCCCTACATT-ATCAGGCCAGAGTGGCTTT-ACAAATCCTC 716
QY 1760 ttaataaaaaataagactcgtcagtaagaagatttt 1800
Db 717 TAAATAACAAATAATAGACTCGCTAGTCAGAACAGATCTCT 757

```

RESULT 11

```

BG747143
LOCUS 602704364F1 NIH_MGC_15 697 bp mRNA linear EST 15-MAY-2001
DEFINITION 602704364F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4857831 5',
mRNA sequence.

```

```

ACCESSION BG747143
VERSION BG747143.1 GI:14057796
KEYWORDS EST.
SOURCE human.

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 697)

```

```

AUTHORS NIH-MGC http://mgi.nhl.gov/.

```

```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

```

```

JOURNAL Unpublished (1999)

```

```

COMMENT Contact: Robert Strausberg, Ph.D.

```

```

Email: cyabbs-remail.nih.gov

```

```

Tissue Procurement: ATCC

```

```

CDNA Library Preparation: Ling Hong/Rubin Laboratory

```

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

```

DNA Sequencing by: NIH Intramural Sequencing Center

```

```

Clone Distribution: MGC clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:

```

```

http://image.llnl.gov

```

```

Plate: L1CM1710 row: o column: 16

```

```

High quality sequence stop: 684.

```

FEATURES

```

source

```

```

1..697

```

```

Location/Qualifiers

```

```

/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"

```

```

/clone="IMAGE:4857831"

```

```

/clone_11b="NIH_MGC_15"

```

```

/tissue_type="adenocarcinoma cell line"

```

```

/lab_host="DH10B (phage-resistant)"

```

```

/notes="Organ: colon; Vector: pOTB7; Site: 1: XhoI; Site_2:

```

```

EcoRI; CDNA made by oligo-dT priming. Directionally

```

```

cloned into EcoRI/XhoI sites using the following 5'

```

```

adapter: GGACACGAG(G). Size-selected >500bp for average

```

```

insert size 1.8kb. Library constructed by Ling Hong in

```

```

the laboratory of Gerald M. Rubin (University of

```

```

California, Berkeley) using ZAP-CDNA synthesis kit

```

```

(Stratagene) and Superscript II RT (Life Technologies)"

```

```

BASE COUNT 177 a 168 g 168 t

```

ORIGIN

```

Query Match 29.3%; Score 679.2; DB 10; Length 697;

```

```

Best Local Similarity 99.4%; Pred. No. 6.5e-167;

```

```

Matches 692; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

```

QY 910 ggaattgtggtttctcttgggacaaatggtctcagaaaattccagagaagaagctatg 969

```

```

Db 2 GGAATTGTGTTTCTCTTGGGATCAATGNTCTCAGAAATTCAGAGAAAGAGCTATG 61

```

```

QY 970 gcaattgctgattgttggcacaataccctcaagacagtcctgtggcggtacactggaa 1029

```

```

Db 62 GCAATTGCTGATGCTTTTGGCAAAATCCCTCAGACAGTCTGTGGCGGTACTGGAA 121

```

```

QY 1030 cgaccatcgaattcttgcgaacaacacagatactgttaagtggctaccccaaaacagactg 1089

```

```

Db 122 CGACCATCGAAATCTTGGCAACACACAGATACTGTTAAGTGGCTACCCCAAAACGATCTG 181

```

```

QY 1090 ctgtgtcaaccgataccctgcttattacccactgattccactggttccatggtttatgaa 1149

```

```

Db 182 CTGTGTACCCGATGACCCGTCCTTTATCACCATGCTGTGTTCCCATGCTGTTATGAA 241

```

```

QY 1150 agcatatgcaatggcgtcccatgggtgatgagtcctgtttgtggtgatcagatggaca 1209

```

```

Db 242 AGCATATGCAATGGCGCTTCCCATGGTGATGTCCTTGTGTGGTATCAGATGGACAAT 301

```

```

QY 1210 gcaagcgcagatgagactaaaggagcgtgagtcacccctgactgttctggaatgactct 1269

```

```

Db 302 CCAAGCGCATGAGACTAAGGAGCTGGAGTGCACCCGTAATGTTCTGGAATGACTTCT 361

```

```

QY 1270 gaagattgaaatgctctaaagcagtcataatgacaaaagttaacagaggaacatc 1329

```

```

Db 362 GAAGATTGAAATGCTCTTAAAGCAGTCATCAATGACAAAAGTTTACAGGAGAACATC 421

```

```

QY 1330 atgcgctctccagccttcacaagaccgcccgggtggagccgctggacccctggcgtgttc 1389

```

```

Db 422 ATGGCGCTCTCCAGCCTTCACAGGACCGCCGGTGGAGCCGCTGGACCTGGCGCTGTC 481

```

```

QY 1390 tgggtggagtgttgatgagcacaaggcggcgccacacctgcccgcagccacacagac 1449

```

```

Db 482 TGGTGGAGTGTGTGATGAGGCAACAGGCGCGCCACACCTTGGCGCCCGCAGGCCAGAG 541

```

```

QY 1450 ctacactgggtacagtcaccattctctggacgtgattggttctcttggcgtgctgtg 1509

```

```

Db 542 CTCACCTGGTACAGTACCATCTCTTGGACGTGATTGTTTCTCTTGGCGCTGCTGCTG 601

```

```

QY 1510 acagtgccttcacacctttaaagtgtgcttattgctaccggaagaaagtctggggaaa 1569

```

```

Db 602 ACAGTGGGCTTCATCACCCTTTAAATGTTGTGCTTATGCTTACCGGAAATGCTTTGGGAAA 661

```

```

QY 1570 aaagggcgagttaaagaagcccaacaaatccaagacc 1605

```

```

Db 662 AAAGGGCGAGTTAAGAAAG-CCACAAATCAAGACC 696

```

RESULT 12

```

AI831562/c

```

```

LOCUS

```

```

DEFINITION

```

```

similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,

```

```

MICROSOMAL (HUMAN);, mRNA sequence.

```

```

AI831562

```

```

wf39b01.xl NCI-CCAP Lu19 Homo sapiens CDNA clone IMAGE:2405161 3'

```

```

764 bp mRNA linear EST 21-DEC-1999

```

```

UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,

```

```

MICROSOMAL (HUMAN);, mRNA sequence.

```



```

QY 683 gtttttcaaaatgccctag--aaatagcctctgaaattctcc-aaacacctgttacgg 739
Db 660 GTTTTTTCAAAAATGCCCTAGGAAATAGCCCTCTGAAATNCTCCAAAACACCTGTACGG 719
QY 740 agtatgatctacagcacacataatttggtgtgtgcgaacgactttgttgact 799
Db 720 AGTATGATCTCTACAGCCACACATCAACTCGGCTCGTGGGACGACGACTGTCGTGACT 779
QY 800 atccaaacccgtgagcccaacatgatcttctcattggttggtatcaac--tgccatcagg 857
Db 780 AT-CCAAACCCGTGATGCCACAATGATCCTCTCTGGGGGTATCCACCTGCCGCTCACGG 838
QY 858 aaagcgttgctatgaatttca 881
Db 839 AAAGCCCTGCTATGGATTGGA 862

RESULT 8
LOCUS AI307269/c
DEFINITION AI307269 818 bp mRNA linear EST 08-APR-1999
similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,
MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION AI307269
VERSION AI307269.1 GI:4002458
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 453.
FEATURES
source
1...818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2055287"
/clone.lib="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/notes="Organ: Kidney; Vector: pMT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 247 a 177 c 134 g 258 t 2 others
ORIGIN

Query Match 30.0%; Score 695.8; DB 9; Length 818;
Best Local Similarity 95.9%; Pred. No. 3e-171;
Matches 745; Conservative 0; Mismatches 29; Indels 3; Gaps 3;

```

```

QY 1547 gctaccggaaatgcttgggggaaaaaggcg- agttaagaaagcccaaatccaaagacc 1605
Db 818 GCTACGCAAAATGCTGGGGGAAAAAGGCGCAATTTAAGNAGGCCCNCAATCCCAAGAAC 759
QY 1606 cattgaagtggttggaataaagtaaaatttgaaacattccctcattcatcaca 1665
Db 758 ATTGGAAGTGGTGGGAATTAAGGTAAATTTGAACCATTCCTCCCAAGTTCATTTCCAA 699
QY 1666 cttgaacaagaataca- gttgtaaatcattttatttatttaagaagaataactttgata 1724
Db 698 CTGAAACAGCAATCAGGGTCAAAATTCATTTATTTCTTATTAAAGGAAATACTTGGCATA 639
QY 1725 aattaatcagcc- acagtgctttaaaaaattctcttaataaaaaataaactagactcgt 1783
Db 638 AATTAATCAGCCCAAGAGTGTCTTTTAAATAATTCCTTAATAAATAAATAAGACTCGCT 579
QY 1784 agtcagaaagataattgaatatgtatogtgcocccctccggtgtcttttgatcagatgac 1843
Db 578 AGTCAGTAAAGATATTTCAATATGTATCGTGCCTCCCTCTGCTTTGATCAGGATGAC 519
QY 1844 atgtccattttcagagcggtgcagacagcgctggcattctagattacttttctactc 1903
Db 518 ATGTGCCATTTTCAGAGGACGTGCAGACAGGCTGGCAATTCAGATTACTTTTCTTACTC 459
QY 1904 tgaacatggtgctgtttgggagtgcggtattcaaaagtggtccacccgctgcccactg 1963
Db 458 TGAACATGCGCTGTTTGGGAGTGGCGGATTCAAAGGTGGTCCACGGTGCCTTACTG 399
QY 1964 caaatgcagtttaacttatcttcttggcttcgcagatggttgcaattgatcttctaac 2023
Db 398 CAATGGCAGTTTAACTTATCTTTGGCTTCGAGATGGTGGCAATTCATTCCTTAAC 339
QY 2024 caataatggtcagtcctcctctctctctctctctctctctctctctctctctctctct 2083
Db 338 CAATAATGTCAGTCCCTCACTCTCTGTCGCTTCATAGTGCACCTCTGTGTGTTAAAG 279
QY 2084 aagggaagcttgctaccttttagatgtagtgaatgaatgaatgagcttgagtgactg 2143
Db 278 AAGGGAAGCTTTTACCTTTAGAGTGTAGTGAATGAATGAATGCTGGAGTGCCTG 219
QY 2144 agaacgcatgattgtttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2203
Db 218 AGAACAGCATATGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 159
QY 2204 gttgatttgagaagataatcattgtcttctcaaatgagctgaatttgataaaaaaccca 2263
Db 158 GTGTATTTCAGAAAGATAATCATTTGCTTATGTCAAATGGAGCTGAATTTGATAAAACCCA 99
QY 2264 aaatacagctatgaagtcgtggcgaagtttcttcttcttcttcttcttcttcttcttct 2320
Db 98 AAATACAGCTATGAAGTCTGGGCAAGTTTACTTTTTTCTGATGTTCTCTACAACT 42

```

RESULT 9

```

LOCUS BG289777
DEFINITION BG289777 881 bp mRNA linear EST 21-FEB-2001
mRNA sequence.
ACCESSION BG289777.1 GI:13045907
VERSION BG289777.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 881)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: AFCC

```

QY	181	gtgtttgtgtcatcgtccagaggttgagtttggccaactgggaagatcactgaatgcaagtg	240
Db	187	GTGCTGTGTAGTTCATGTCGACAGAGGTGAGTTTGGCACTGGGAAGATCAGTGAATTCACAGCTG	246
QY	241	aagacttatccaacttcataataccctgagagatctggaccggaggttcaagagcttttggcc	300
Db	247	AAGNCATTATCACTTTCATATACCTCTGGAGGATCTGGACCGGAGTTTCAAGGCTTTTGCC	306
QY	301	catgtcctaattggaagcaaatgacyaagtatatattctctattaatgggtttcacacaat	360
Db	307	CATGCTCAATGGAAAGCACAACTAGCAAGATATATATCTCTATTAAATGGTGTCAACAAT	366
QY	361	gacatttttgactta-tttttttcaaatgcgaagdtgttttaaagacaaaataatgatt	419
Db	367	GACATTTTGTACTTATTTTTCAAATTCAGAGAGTTGTGTTTAAAGACAAAAATTAGT	426
QY	420	agaactacttaagaagagttc-ttttgatgcaggtttctcgatcccttttgataactgtg	478
Db	427	AGAACTACTTAAGAGAGAGTTCTTTTGTATGACGTGTTCGTGATCTCTTTGATAAAGTGTG	486
QY	479	gcttaattgttgcacaatatcttccctccctccgtggttctgcgcaggggaatacttt	538
Db	487	GCTTAAATGTTGCCAAATATTTCTCCCTCCCTCCGTGCTTCGCGACGGCGAATACTTT	546
QY	539	gccctctatcttgaagaagtgcaagtgccctgctcctcttctctatgtccccaagaattc	598
Db	547	GCCACTATCTTGAGAGAGTGCACATGCCCTGCTCTCTTCTATGTGCCCGAAGATTC	606
QY	599	tcttaaggttctcagatgcatttcgaagtgagtgactcaaggaagaccac-atcatgcac	657
Db	607	TCCTTAGGTTCTCAGATGCCATGACTTCAAGGAGAGAGTAGGAACCAACCAATCATGCAC	666
QY	658	ttggaggaacatttattatgccaccgttttttcaaaaatgccttagaataagctcttgaa	717
Db	667	TTGEGGAACATTTATTATGCCACCGATTTTCAAAAATGCCCTAGAAATAGGCTCTGAA	726
QY	718	attcttccaaacacacttt-acagagtatga- ---tctctacagccacacatacaattggtt	773
Db	727	ATTCTCCAAAGACCTGTTAAACGGAGTATGAATCTCTACAGACACACACTCAAAATGGAAAT	786
QY	774	gttgcgaacgg--actttgttttggactatcccaaacccgt-gatgcccaacatgatctt	830
Db	787	GTTCGGAACGGGACCTTTGTTGGGAACATATCCAAACCCGTGGATGCCCAACATGATCTT	846
QY	831	catgttggttatcaactgcatacgaaggaaa	860
Db	847	CATGGGAAAAACAACGTGTATTTCGTGTAAAT	876

RESULT	7
BI762347	
LOCUS	
DEFINITION	BI762347 884 bp mRNA linear EST 25-SEP-2001
	603049004F1 NTH_MGC_116 Homo sapiens cDNA clone IMAGE:5189040 5', mRNA sequence.

SOURCE Illumina.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 884)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

found through the I.M.A.G.E. Consortium/LLML at:
http://image.llnl.gov
Plate: LLML1472 row: h column: 01
High quality sequence stop: 796.
Location/Qualifiers
1. .884
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5189040"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/notes="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
213 a 207 c 210 g 253 t 1 others
BASE COUNT
ORIGIN

```

```

BASE COUNT      213 a      207 c      210 g      253 t      1 others
ORIGIN

Query Match      30.5%;      Score 707.4;      DB 10;      Length 884;
Best Local Similarity 94.0%;      Pred. No. 2.8e-174;
Matches 812;      Conservative 0;      Mismatches 42;      Indels 10;      Gaps 7

```


found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10407 row: m column: 12

High quality sequence stop: 746.

FEATURES

Location/Qualifiers

1..882

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4516403"

/clone_lib="NIH_MGC_93"

/tissue_type="transitional cell papilloma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: bladder; Vector: pCMV-Sport6; Site:1: NotI;

Site:2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

259 a 155 c 202 g 266 t

BASE COUNT

ORIGIN

Query Match 31.18; Score 722.4; DB 10; Length 882;

Best Local Similarity 94.5%; Pred. No. 3.3e-178;

Matches 792; Conservative 0; Mismatches 41; Indels 5; Gaps 4;

QY 1447 gacccacccgtgaccagaccattcctggacgtgattggttctctcttgcgcgcgtg 1506

|||||

Db 1 GACCTCAGCTGGTACCAGTACCATTCCTGGACGTGATTGGTTCTCTTGGCCGTCGTG 60

QY 1507 ctgacatgacctcattcaccatttaattgtcttgcctaccggaatgcttgagg 1566

|||||

Db 61 CTGACATGGCCCTCATCACCCTTAAATGTTGTGTGTTAGCTACCGGAATGCTGGG 120

QY 1567 aaaaaggcgaggttaagaagcccaaatccaaagaccattgagaagtgggtggaaa 1626

|||||

Db 121 AAAAAAGGCGAGTTAAGAAAGCCCAAAATCCAGACCCATTGAGAAGTGGTGGAAA 180

QY 1627 taaggtaaatttgaacattccctagtcatttccaaacttgaacacagaatcagttt 1686

|||||

Db 181 TAAGGTAAATTTGNAACCATTCCTAGTCAATTCCTCAACATGAAACAGAAATCAGTGT 240

QY 1687 aaattcattttatttatttaagaataacttttgataataataatcagcccgagtgct 1746

|||||

Db 241 AAATTCATTTTATTCTTATTAAAGAAATACTTTGCATAAATAATTAATCAGCCAGAGTCT 300

QY 1747 taaagaattctttaaataaaataaataagactcgtagtcagtaagaatatttgaatat 1806

|||||

Db 301 TTAATAATTCCTTAAATAAATAATAGACTCGCTAGTCACTAAGATATTGAATAT 360

QY 1807 gtatgccccctccggtgtcttggatcaggaatgacatgtgccattttccagagcgt 1866

|||||

Db 361 GTATCGTCCCGCTCCGGTGTCTTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAGCT 420

QY 1867 gagacaggtcgacattctagatttcttcttactctgaacatggcctgttggaggt 1926

|||||

Db 421 GCAGACAGGCTGGCATCTAGATTACTTTTCTTACTCTGAACATGGCCTGTTGGAGT 480

QY 1927 gcgggattcaagggtgtccaccgctgccccctactgcaaatggcaggttttaatttacc 1986

|||||

Db 481 GCGGGATTCAAAAGGTGTGCCACCGCTGCCCTTACTGCAAAATGGCAGTTTAACTTATC 540

QY 1987 tttagctcttcagatggttgcaattgattcattcaacaaatggcagtcctcatctc 2046

|||||

Db 541 TTTTGGCTCTCGAGATGTTGCAATTTGATTCCTTAACCAATAATAGTTCAGTCTCATCTC 600

QY 2047 tgcctccttcattagtgccacctgtgtt-gtttaagaagggaagctttgtacctttag 2105

|||||

Db 601 TGTCCTGCTTCATAGTGCACACTTGTGTGGTTTAAAGAAGGAAGCGTTGTACCTTTAG 660

QY 2106 agtcagggtg-aaatgaatgaatggcttgagtgacagagacagcatatgatcttct-t 2163

|||||

Db 661 ACTGTAGGTGAAATGAATGAATGTTGGAGGTGCACATGAGAACAGCATATGATTCTGG 720

QY 2164 gctttgggaaagaatgctgctgataaattggtgggtgtgtgtatttgagaataatc 2223

|||||

Db 721 GCTTTGGGAAAAAGAAATGATGCTATGGAATGGGGGTGGTGGAGAAAAATATC 780

QY 2224 attgcttatgcaaatgagctgaatttgataaaaccccaatacacagctatgaagt 2281

|||||

Db 781 ATGCTTTATGTCNAAGAGG--TGATTTGTAATAAACCAATAACAGGTATTAAGGTG 836

RESULT

6

BI759297

LOCUS

603042939Pl NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5183433 5',

DEFINITION

mRNA sequence.

ACCESSION

BI759297

VERSION

BI759297.1 GI:15750875

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 876)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue procurement: Life Technologies, Inc.

cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11457 row: n column: 10

High quality sequence stop: 735.

Location/Qualifiers

1..876

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5183433"

/clone_lib="NIH_MGC_116"

/lab_host="DH10B"

/note="Organ: pooled colon, kidney, stomach; Vector:

pCMV-Sport6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA

source anonymous pool of 3 colons, age 26 yo male, 49 yo

female, 71 yo male colon; 46 yo male kidney, and pool of 2

stomachs, 62 yo male and 70 yo female. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.4 kb,

insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C.

Gruber (Invitrogen). Research Genetics tracking code

023. Note: this is a NIH_MGC Library."

226 a 193 c 205 g 252 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 94.6%;

Matches 823; Conservative 0; Mismatches 37; Indels 10; Gaps 7;

QY 1 ctcagctgcagttctctgattgctgacaggggtggaccagccctctctctatgtg 60

|||||

Db 7 CTCAGCTGCAGTCTCTGATGGCTTGACAGGGTGGACCAGCCCTCTCTATGTGTG 66

QY 61 tgcctgctgctgacctgtgctttgcagagcaggggaagctactgtagtgcctatgat 120

|||||

Db 67 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGAGGCAAGCTACTGCTAGTGGCCATGAT 126

QY 121 gggagccactgggttcaccatgagtcggtgggtggagaaactcattctcagggggcagag 180

|||||

Db 127 GGGAGCCACTGGTTCCACCATGAGTGGTGGTGGAGAACTCATTTCTCAGGGGGCATGAG 186


```

source      1. .2450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3854370"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
BASE COUNT 669 a 509 c 565 g 707 t
ORIGIN

Query Match      70.3%   Score 1631.6;   DB 11;   Length 2450;
Best Local Similarity 82.6%   Pred. No. 0;
Matches 1880; Conservative 0; Mismatches 394; Indels 2; Gaps 1;

Qy 47 ttcctctatgtgtctctgtctgtacctgtgtcttgcgcagcgaggaagactactgg 106
Db 128 TTTCGCAAGGGTTTCTCTTTAGCACCTTGGGGCATGTTGTAGGTGACAAAGCTCTGG 187

Qy 107 tegtgtccatgtatggagccactgtttcacccatgaggtcggtgtgtgtgagaaactcattc 166
Db 188 TGGTCCCTCAGGACGGAAGCCACTGGCTTATGATGAAGGATATAGTTGAGGTTCTCAGTG 247

Qy 167 tcaggggcatgaggtgtgtgtatgcatgcagagtgagttgtgcaactgggaagatac 226
Db 248 ACCGGGGTCATGAGATGTAGTGTGGTGGCTGAAATTAAATTTGTTTGAAGAATCCA 307

Qy 227 tgaattgcacagtgagactattcaacttcataaccctggagagatctgtgacccggagt 286
Db 308 AATACTACACAAGAAAATCTATCCAGTGGCTATGACCAAGAGAGCTGAAGAACCGTT 367

Qy 287 tcaaggtctttgtccatgtctcaatgg--aaagcaagtaacgaagtatatctctatt 344
Db 368 ACCATCATTTGGAAACAATCTTGTGAGCGATCATCTCTAACTGCTCCTCAGACAGA 427

Qy 345 aatgggttcacaaatgacatttttgactatttttttcaaatgtcagggatttggtaa 404
Db 428 GTACAGGAATAACATGATTGTTATTTGGCTGTACTTCACTCAACTGCCAGAGCCTCTGCA 487

Qy 405 aacacaaaaatagtagaactataaaggagagttcttttgatcagtggtttctcgatcc 464
Db 488 GGACAGGACACCCCTGAACCTTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACC 547

Qy 465 ttttgataactgtgtgttaattgttgccaaatatttctccctccctccgtgtgtcttcgc 524
Db 548 AGCCTTACCCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCTTACCCTGTGTACCTCTT 607

Qy 525 cagggaataactttgtccactacttgaagaaggtgcaagtgccctgtctctctctccta 584
Db 608 CAGGGGTTTTCCGTGTCCTGGAGCATPACATTCAGCAGAAAGCCCAACCTGTGTCTTA 667

Qy 585 tgcctccagaattctctgttctcagatgaccttcaaggagagagtaacggaa 644
Db 668 CATTCGCCAGGTGTACACAAAATTTTTCAGACCAACATGACTTTTTTCCCAACGAGTGGCCAA 727

Qy 645 ccacatcgtacacttgaggaacatttatatgccaccgttttttcaaaaatgcctaga 704
Db 728 CTTCTCTTTAAATTTGTTGGACCTATCTATTTTATTTGTTCTGTTTTCAAAGTATGAAGA 787

Qy 705 aatagcctctgaaattctccaaacacctgttcaaggagtagtatctctacagaccacatc 764
Db 788 ACTGCATCAGCTGCTCCTCAAGAGAGATGTGGATATAATCACTTTATATCAGAAAGGTCTC 847

Qy 765 aatttggtgtgtgcagcagcactttgttgactatcccaaacccgtgtatgccaact 824
Db 848 TGTGTGGCTGTTAAGATATGACTTTGTGCTTGAATATCTAGTGGCGGTCTATGCCCAACAT 907

Qy 825 gatctctatgtgtgtgtatcaactgcacatcaagggaagccgtttgctctatggaaattgaagc 884
Db 908 GGTCTTCATTTGGAGGTATCAACTGTGAAGAAGAGGAAGAACTTGTCTCAGGAAATTTGAAGC 967

Qy 885 ctacatataatgtctctctggagacatggaattgtgtgtttctctcttttgggatacaatgggtctc 944

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 02:43:12 ; Search time 4821.79 Seconds
(without alignments)
6494.048 Million cell updates/sec

Title: US-09-835-082-1
Perfect score: 2320

Sequence: 1 ctacgctgcagttctctgat.....ttctgatgtttctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1631.6	70.3	2450	11	BC011409 Homo sapi
2	789.6	34.0	2306	11	BC012716 Mus muscu
3	780.2	33.6	2202	11	AK002629 Mus muscu
4	741.2	31.9	868	10	BC402908 602418710
5	722.4	31.1	882	10	BC291462 602387347
6	720.8	31.1	876	10	BI759297 603042939
7	707.4	30.5	884	10	BI762347 603049004
8	695.8	30.0	818	9	AI307269 tb24dl2.x
9	695.4	30.0	881	10	BC289777 602385071
10	680.6	29.3	1005	10	BE870911 604448662
11	679.2	28.3	697	10	BE747143 602704364
12	669.4	28.9	764	9	AI831562 602447772
13	644	27.8	723	10	BC424452 602447772
14	641.2	27.6	829	10	BC469506 602532818
15	640	27.6	833	10	BC400591 602464244
16	639.8	27.6	791	9	AI765716 w183f06.x
17	636.8	27.4	658	9	AV652040 AV652040

18	635.4	27.4	945	10	BC167885
19	625.6	27.0	724	10	BF031444
20	621.8	26.8	688	10	BC286763
21	618.8	26.7	796	10	BC432363
c 22	604.8	26.1	861	9	AI829268
23	601	25.9	647	9	AV692287
24	596	25.7	763	10	BC399019
25	594.2	25.6	663	9	AV646537
26	592.4	25.5	859	10	BE249993
c 27	589.8	25.4	640	9	AI635995
c 28	588	25.3	683	9	AI480360
29	584.8	25.2	812	10	BC401258
c 30	584	25.2	621	9	AW151709
c 31	583	25.1	667	9	AW051715
32	583	25.1	969	10	BC291839
c 33	572	24.7	636	9	AI767470
34	567.4	24.5	581	10	BE379923
35	567	24.4	623	10	BC171502
36	557.8	24.0	743	10	BC386668
37	556.4	24.0	741	10	BC433491
c 38	550.6	23.7	619	9	AW300312
39	549.8	23.7	851	10	BC401340
40	543.8	23.4	700	10	BF130433
c 41	542.8	23.4	608	9	AW469219
42	542.8	23.4	953	10	BC328092
43	537	23.1	935	10	BF234991
c 44	535.4	23.1	626	9	AW082933
45	532.2	22.9	666	10	BF027719

ALIGNMENTS

RESULT	1
BC011409	
LOCUS	BC011409
DEFINITION	Homo sapiens, clone IMAGE:3854370, mRNA.
ACCESSION	BC011409
VERSION	BC011409.1 GI:15030273
KEYWORDS	HTC.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 2450)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL) Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: villalon@bcm.tmc.edu Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 13 Row: b Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4507814
This clone has the following problem: incomplete processing.
Location/Qualifiers

QY 1459 taccagtaccattccttgacgtgattggtttctcttcttgccgtgtgtgacagtgcc 1518
|||||
Db 1441 TACCAGTACCATTCTTGACGTGATTGGTTTCTCTTGCCCGTGTGACAGTGCC 1500
|||||
QY 1519 ttcacacccctttaaatgtgtgcttatggtaccggaatgcttggtgggaaaaagcgca 1578
|||||
Db 1501 TTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGAAAAAGGCGA 1560
|||||
QY 1579 gttagaagaagccacaaatccaagaccattga 1611
|||||
Db 1561 GTTAGAAAAGCCCAAAATCCAGACCCATTGA 1593
|||||

Search completed: August 21, 2002, 11:58:40
Job time: 32903 sec

Db 1256 CCTTGTGTTGATCAGATGGCAATGCAAGCCGATGGAGACTAAGGAGCTGGAGTGA 1315
 QY 1244 cccgaattcttggaataaagctctgaagatttagaaaaatgctctctaaagcagctatca 1303
 Db 1316 CCCGTAATGTTCTGGAATGACTTCTGAGATTTAGAAAATGCTTAAAGACGATCATCA 1375
 QY 1304 atgacaaaagttaacaaggagaacatcatgcctctccagctctcaagaccgcccgg 1363
 Db 1376 ATGACAAAAGTTACAAGAGAACATCATGGCCCTCTCCAGCCCTTCAAGGACCCGCCGG 1435
 QY 1364 tggagccgctggagactggccgctgtctgtggtgagttgtgtagaggcacaagggcgcc 1423
 Db 1436 TGGAGCCGCTGGACCTGGCCGCTGCTGGGTGGAGTTGTGATGAGGCACAAAGGGCGCC 1495
 QY 1424 cacactctgcccgcagccagccacccactcactgtgtaccagaccattcccttgagtgga 1483
 Db 1496 CACACTGGCCCGCCGAGCCAGCCAGCCCTCAGCTGTACACAGTACCATCTCTTGAGCTGA 1555
 QY 1484 ttggtttctcttgccgctgctgctgacagtggcctctccatcccttttaattgtgtgctt 1543
 Db 1556 TTGGTTTCTCTTGGCCGCTGCTGACAGTGGCCTTCAATCACTTTAAATGTTGTCTT 1615
 QY 1544 atggctaccggaatgtgtgggaaagggcaggttaagaaagcccaaatccaaag 1603
 Db 1616 ATGGCTACCCGAATGCTTGGGGAAGAAAGGCGAGTTAAGAAAGCCACCAATCCAAAG 1675
 QY 1604 cccattgagaagtggttggaataaaggttaaaattttgaacccatccctgacttccaa 1663
 Db 1676 CCATGTGAGAGTGGGTGGGAATTAAGGTAAATTTTGAACCA-TCCCTAGTCAATTCGA 1734
 QY 1664 aacttgaacacagaaatcagtggttaattcattcttcttatttataagaaatcttgcat 1723
 Db 1735 AACTTGAACACAGAAATCAGTGTAAATTCATTTATTCATTAAGGAAATACCTTTGCA 1794
 QY 1724 aaattaacagccccagagt-g-ctttaaaaaattctcttaataaaaaataagactcgc 1782
 Db 1795 AAATTAATCAGCCCGAGAGTGCTTTTAAATAATCTCTTAAATAAAATAAAGACTCGC 1854
 QY 1783 tagtcagtaaaagatttgaatgatgatcgcccccctccgctgctcttgcagagata 1842
 Db 1855 TAGTCAGTAAAGATATTGAATATGATATGTCGTGCCCCCTCGGAGTCTTTGATCAGATGA 1914
 QY 1843 catgtgcaatttttcagagaagctgcagacagagctggaattctagatttaattcttact 1902
 Db 1915 CATGTGCCATTTTTCAGAGAGCGTGCAGACAGCTGGCATTCTAGATTACTTTTCTTACT 1974
 QY 1903 ctgaacacatggcctgtttggagtgcgggattcaaaaggtgggtccacgcgtgccctact 1962
 Db 1975 CTGAACATGGCCTGTTGGGAGTGGGGATTCAAAAGGTGGTCCACGGCTGCCCTACT 2034
 QY 1963 gcaaatgagcttttactcttcttcttcttcttcttcttcttcttcttcttcttcttct 2022
 Db 2035 GCATATGGCAGTATTAATCTTATCTTTTGGCTCTTCAGATGGTTCG-ATTGATCTCTAA 2093
 QY 2023 ccaaataatgctcctcctctctctctctctctctctctctctctctctctctctctct 2081
 Db 2094 CCAATATGTTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2152
 QY 2082 aaaaaggaagcttttaccctttagagtgtagtgaaatgaatgaatgaatgaatgaatgaat 2141
 Db 2153 AGAAGGGAAGCTTTGATACCTTTAGAGTGTAGTGAATGAATGAATGAATGAATGAATGA 2212
 QY 2142 taagaaacagataatattcttcttcttcttcttcttcttcttcttcttcttcttct 2201
 Db 2213 TGAGAACAGGATATGATTTCTTGTGTTGGGGAAGAAAGAAATGATGATGAATTTGTGGG 2272
 QY 2202 tggttatttgaagaataaataatcttcttcttcttcttcttcttcttcttcttcttct 2261
 Db 2273 TGGTGTATTTGAGAGATAATCATGCTTATGTCATTAATGGAGCTGAATTTGATAAAACC 2332
 QY 2262 caaantacagctatgaagtgcctgggcaagttac-ttttttctgtatcttctctcaact 2320
 Db 2333 CAAATACAGCTATGAAGTGTGGGCAAGTTTACTTTTTTCTGATGTTTCTCTCAACT 2392

RESULT 14

HUMUGT1FA

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

gene

CDS

polyA_signal

polyA_signal

polyA_site

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 1870;

Conservative

0;

Mismatch

398;

Indels

12;

Gaps

8;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

HUMUGT1FA 2422 bp mRNA linear PRI 14-JAN-1995
 Homo sapiens phenol UDP-glucuronosyltransferase (UDPGT) mRNA,
 complete cds.

J04093.1 GI:624724
 UDP-glucuronosyltransferase; phenol.
 Homo sapiens kidney cDNA to mRNA.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (Bases 1 to 2422)
 Harding, D., Fournel-Gigleux, S., Jackson, M.R. and Burchell, B.
 Cloning and substrate specificity of a human phenol
 UDP-glucuronosyltransferase expressed in COS-7 cells
 Proc. Natl. Acad. Sci. U.S.A. 85 (22), 8381-8385 (1988)

89042187
 On Jan 18, 1995 this sequence version replaced gi:340517.
 Location/Qualifiers

1..2422
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="kidney"
 1..2422
 /gene="UDPGT"
 88..1683

/gene="UDPGT"
 /EC_number="2.4.1.17"
 /codon_start=1
 /product="UDP-glucuronosyltransferase"
 /protein_id="AAA61251.1"
 /db_xref="GI:624725"

/translation="MACLLRSQRLSAGVFELALWGVYCDKLLVVPQDSHWLSMKD
 LIVEVLSDRGHEIVVPEVNNLLKEYKTYTRKTYPPYDDEELKNRYQSGNNHFAER
 SELTAPQTEYRNNMIVIGLFFINCSLLQDRDTUNFKESKEDALFTDPALPCGVILA
 EILGFLPSYLFKRGFPSCLEHTFSRDPVSPVYPCRYKFSDHMTSRVANFLVLLIE
 PLYFLCSKYKELASAVLKRDVDTITLSEVSWLLRYDFVLEPRPYNNMVFIGGI
 NCKRKDLQSDEFEAYINASGHEIVFSGVSWSEIPEKKAMAIADALGKNPQTVLWR
 YTKTRPSNLANTILVKWLPPNDLLGHPMTAFITTAGSHGVETISICNGVPMVMPLE
 GQMDNAKRMETKAGVTLNLEMTSEDLNALKAVINDKSYKENTIMRLSSLSLHDKRPV
 EPDLIAVFWVEYVNRHKGAPHLPAHAHLDTWYQVHSLDVGFLAVLVITAFITFKCC
 PYGYPKQJLGKKGKVRKAHKSHTH"

2396..2401
 /gene="UDPGT"
 2404..2409
 /gene="UDPGT"
 2422

/gene="UDPGT"
 653 a 516 c 555 g 698 t

polyA_signal
 polyA_signal
 polyA_site

BASE COUNT 653 a 516 c 555 g 698 t
 ORIGIN

Query Match 66.3%; Score 1539.2; DB 9; Length 2422;
 Best Local Similarity 82.0%; Pred. NO. 0;
 Matches 1870; Conservative 0; Mismatches 398; Indels 12; Gaps 8;

QY 47 ttccctctatgtgtgtctgtctgtgacctgtgcttccagagggcagggagactactgg 106
 Db 119 TTTCTGCAGGGGTTTCTTCTTCTTAGCTCTTTGGGCGATGTTGTAGGTGACAGCTCTGG 178

QY 107 tagtgcacagagatggagccacactggttccactaggtcggtgtggtggagaaactcattc 166
 Db 179 TGGTCCCTCAGGAGCGAGCCACCTGGCTTAGTATGAAGATATAGTTGAGGTTCTCAGTG 238

QY 167 tcagggggcagtaggtgtgtgttagtcatgcacagaggtgagttggcaactgggaagatcac 226
 Db 239 ACCGGGGTCTAGATAGTTGTAGTGTGTGCTGCTCAAGTTAAATTTCTTTTGAAGATACA 298

QY 227 tgaattgcacagtagaagactattcaactctcataccctggagagactctggaccggggagt 286

QY 1805 atgtatcgtgccccctccgtgtcttcttgatcagatgacatgycatttcttcaggagac 1864
 DB 1811 ATGTATCTGTCGCCCCCTCCGCTGCTTTGATCAGGATGACATGTGCGCATTTTTCAGAGGAC 1870
 QY 1865 gtccacacagctgacattctagattactttctactctgaaacatgacctgtttggga 1924
 DB 1871 GTCCAGACAGGCTGCGATCTAGATTACTTTCTTACTCTGAAACATGCGCTGTTGGGA 1930
 QY 1925 gtccggattccaaagggtgtccacccgctgcccctactgcaaatggcagttttaetcta 1984
 DB 1931 GTGCGGATTCAAAAGGTGGTCCACCGCTGCCCTACTGCAAAATGGCAGTTTAACTTA 1990
 QY 1985 tctttggctctccagatgttgcaattgaccttaacacataatgtcagtcctcattc 2044
 DB 1991 TCCTTTGGCTCTCGCATGTTGCAATGATGATCTTAAACCAATATGTTGTCAGTCTCTC 2050
 QY 2045 tctgtcctctctcattagtgccacacctgtgtgttttaagaagggaagctttgtacettta 2104
 DB 2051 TCTGTCTCTCTTCAATAGGTGCCACCTTGTGTGTTTAAAGGAAGGAAGCTTTGTACCTTA 2110
 QY 2105 gagttagtgaaatgaatgctgtgagtgacactgagacagcatatgatttcttg 2164
 DB 2111 GAGTGTAGTGAATGAATGAATGGCTTGGAGTGCACCTGAGACACATATGATTCTTG 2170
 QY 2165 ctttggggaaagaatgactgactgaaattgggtgggtgttatttgagaagaataca 2224
 DB 2171 CTTTGGGGAAAAAAGATGATGCTATGAAATTGGTGGTGTGTTTGAAGAAGATAATCA 2230
 QY 2225 ttgcttatgcacaaatggagctgaattgataaaaccccaaaatacagctatgaagtgctg 2284
 DB 2231 TTGCTTATGTCAAATGGAGCTGAATTTGTATTAACCAACCAAAATACAGCTATGAAGTCTG 2290
 QY 2285 ggcgaagttaactttttcttgatgttcttccataact 2320
 DB 2291 GGCAAGTTTACTTTTTTCTGATGTTTCTCTACAACT 2326

RESULT 13
 LOCUS AX336711 2422 bp DNA linear PAT 09-JAN-2002
 DEFINITION Sequence 7220 from Patent WO0194629.
 ACCESSION AX336711
 VERSION AX336711.1 GI:18127430
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)
 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
 Horrigan, S., Soppet, D.R. and Weaver, Z.
 Cancer gene determination and therapeutic screening using signature
 gene sets
 JOURNAL Patent: WO 0194629-A 7220 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES Location/Qualifiers
 Source 1..2422
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 653 a 516 c 555 g 698 t
 ORIGIN

Query Match 66.3%; Score 1539.2; DB 6; Length 2422;
 Best Local Similarity 82.08; Pred. No. 0;
 Matches 1870; Conservative 0; Mismatches 398; Indels 12; Gaps 8;

QY 47 ttcctctatgtgtgtctgtcgtgacacctgtgcttggccaggcagggaagctactg 106
 DB 119 TTTCTGACGGGGTTTCTTCTTACTGCTTTTGGGGCATGGTGTAGGTGACAAAGCTGTGG 178
 QY 107 tagtgcccatggatggggagccactggttccaccatgaggtcgggtggggagaactcattc 166

DB 179 TGTCTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGATATAGTTGAGTCTCAGTG 238
 QY 167 tcaaggggcatagtggttgtagtcatgccagaggttgagttggcacaactgggaagatcac 226
 DB 239 ACCGGGTCATCAGATCTAGTGGTGGCTGAAGTTAATTTGCTTTTGAAGAATAACA 298
 QY 227 tgaattgcacagtgaaagacttattcaacttcataataccctggagatctggacggaggt 286
 DB 299 AATACTACACAGAAATAATCTATCCAGTGGCGTATGACCAGAGAGCTGAAGAACCGTT 358
 QY 287 tcaagggcttttggccatgctcgaatggaaagcacaagtacgaagatatattctctattaa 346
 DB 359 ACCAATCATTTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCTCAGACAG 418
 QY 347 tgggttcatacaatgaca---tttttgacttattttttcaaaatgcaggagttgttta 403
 DB 419 AGTACAGGAATAACATGATTGTTATTTGGCCTGTACTTCATCAACTGCCAGAGCCTCTCGC 478
 QY 404 aagacacaaaaatagtagaataactaaaggagagttcttctttgatgcagtttctcgatc 463
 DB 479 AGGACAGGGACACCTCTCAACTTCTTTAAGGAGAGCAAGTTGATGCTCTTTTCAGAGCC 538
 QY 464 cttttgataactgtgcttaattgttgccaaaattttccctccctccctccgtgcttg 523
 DB 539 CAGCCTTACCTGTGGGGTATCCTGGCTGAGTATTTTGGCCCTACCATCTGTGTACCTCT 598
 QY 524 ccaggggaataacttggccactatcttgaagaaggtgcacagtgccctgctcctcttctcct 583
 DB 599 TCAGGGGTTTCCGTGTCTCGAGCATACATTCACAGAGAGCCAGACCTCTGTCTCT 658
 QY 584 atgtccccagaattctttaggttctcagatgcccactgttcaagagagagatgcagg 643
 DB 659 ACATCCCAAGTGTACACAAAGTTTTCAGACCACATGACTTTTCCCAACGAGTGGCCA 718
 QY 644 accacatcatgcacttggaggacaatttattatgccacogtttttccaaaatgccctag 703
 DB 719 ACTTCTTGTAAATTTGTTGGAGCCCTATCTATTTTATTGCTGTGTTTCAAGATATGAGA 778
 QY 704 aaatagcctctgaattctccaaacacctgttcaggagatgatctctacagccacacat 763
 DB 779 AACTCGCATCAGCTGTCTCAAGAGAGATGT---GGATATAATCACTTATCAGAGTCT 835
 QY 764 caatttgggttgtgcgaacggaccttggtttggactatcccaaacccgtgagtcacca 823
 DB 836 CTGTTTGGCTGTAAAGATATGACTTTGTGCTTGAATATCTAGGCGGCTATGCCCAACA 895
 QY 824 tgatctcattgggtgtatcaactgcatcagggaaagccgttgcttatggaatttgaag 883
 DB 896 TGTCTTCAATGGAGGTATCACTGTAAAGAGAGAAAGACTTGTCTCAGGAATTTTGAAG 955
 QY 884 cctacattatgcttctggagaacatggcaattgtgttttctctcttggggatcaatggct 943
 DB 956 CCTACATTAAATGCTCTCGAGAACATGGAATTTGTTGTTTCTCTTTGGGATCAATGTCT 1015
 QY 944 cagaaattccagagaagaagctatggcaattgtctgactgtgttgggcaaaatccctcaga 1003
 DB 1016 CAGAAATTCAGAGAGAAAGCTATGGCAATTTGCTGATGCTTTTGGGCAAAACCTCTCAGA 1075
 QY 1004 cagtccctgagcggtacactggaaaccccgaccatggaattcttgggaacacacagctactg 1063
 DB 1076 CAGTCTGTGGCGGTACACTTGGAAACCCGACCATCGAATCTTTGGCAACACACGATACTTG 1135
 QY 1064 ttaagtggttaccaccccaaacgactctgttgggtcaccggatgacccgtgcttttaccacc 1123
 DB 1136 TTAAGTGGCTACCCCAAAACGATCTGCTTGTGTCACCCGATGACCGTGTCTTTATCACCC 1195
 QY 1124 atgctgggttcccatggtgtttatgaagcatatgcaatggcggttcccatggtgatgac 1183
 DB 1196 ATGCTGGTTCCTCATGGTGTGTTTGAAGCATATGCAATGGCGTTCCTCATGGTATGATGC 1255
 QY 1184 ccttggttgtgatacagatggacaatgcaacgcgcatggagactaaaggagctggagtga 1243


```

Db 726 AACTCCGATCAGCTGCTCTCAAGAGAGATGCGATATAATCACTTATATCAGAAAGTCT 785
Qy 764 caatttggttgtagaacgaccttggttttgtagactatcccaaacccgtagtgcocaa 823
Db 786 CTGTTTGGCTGTTAGATATGACTTGTGCTTGATATCTAGCCGGTCTAGCCCAACA 845
Qy 824 tgatcttoatttggttgtagaacgaccttggttttgtagactatcccaaacccgtagtgcocaa 883
Db 846 TGGTCTTCATTCGAGGATATCAACTGTAAAGAGAGAAAGACTTGTCTCAGGAATTTGAAG 905
Qy 884 cctacataatgctcttgtagaacgaccttggttttgtagactatcccaaacccgtagtgcocaa 943
Db 906 CTTACATTAATCTCTTGAGAACATGGAATGTTGTTTCTCTTTGGGATCAATGCTCT 965
Qy 944 cagaaatccagagaagaagctatggcaattgctgatgctttgggcaaaatccctcaga 1003
Db 966 CAGAAATTCAGAGAGAAAGACTATGCAACTGCTGATGCTTTGGGCAAAATCCCTCAGA 1025
Qy 1004 cagtcctgtggtgtagaacgaccttggttttgtagactatcccaaacccgtagtgcocaa 1063
Db 1026 CAGTCTGTGGCGGTACACTGGAAACCCGACCATCGAATCTTGGCAACACAGATCTG 1085
Qy 1064 ttaagtggctaccccacaaacgactctggttcacccgtagtgcocaaatcccttatacc 1123
Db 1086 TTAAGTGGCTACCCCAACAGATCTGCTTTGGTCAACCGGATGACCCGCTTTATCACCC 1145
Qy 1124 atgctggttccatggtttatgaaagcatatgcaatggcgttccatggtgatgac 1183
Db 1146 ATGCTGTTCCCATCGGTGTTATGAAGGATATGCAATGCGGTTCCCATGTTGATGTC 1205
Qy 1184 cctgttggtgatcagatggacaatgcaagcgcatggagactaaggagctgagtgatga 1243
Db 1206 CTTGTTGTTGATCAGATGACAAATGCAAAAGCGCATGGAGACTAAGGAGCTGAGTGA 1265
Qy 1244 ccttgatgtctggaatgactctggaagatttagaaatgctctaaagcagtcacatca 1303
Db 1266 CCTGATGTTCTGGAATGACTCTGGAATTTAGAAATGCTCTAAAGCAGTATCA 1325
Qy 1304 atgacaaagtataaggagaaacatcatgctgctctccagcttcacaaaggacgcccgg 1363
Db 1326 ATGACAAAGTTTACAAAGAGAACATCATGCGCTCTCCAGCTTCACAAAGACCGCCCGG 1385
Qy 1364 tggagcgcgtgacccgctgctctggtggtgagttgtgtagagcacaaggcgccgc 1423
Db 1386 TGGAGCGCTGACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1445
Qy 1424 cacaactgccccagccagccagccacactcactgtagtaccatcttccttgacgtga 1483
Db 1446 CACACCTGCGCCCGCCAGCCACGACCTCCTGTTACCAAGTACCATCTCTTGACGTGA 1505
Qy 1484 ttggttctcttggcgcgtgctgacagtgcccttcacacccctttaaattgttgctt 1543
Db 1506 TTGCTTTCTCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1565
Qy 1544 atgctacagaaatgcttgaggaaagagccaggttaagaaagccacaaatccaaaga 1603
Db 1566 ATGCTACCGGAAATGCTTGGGAAAGAGGCGAGTTAAGAAAGCCCAATCCCAAGA 1625
Qy 1604 cccattgagaagtgggtgggaataaggttaaaatttgaaccattcccttagtcattcca 1663
Db 1626 CCCATTGAGAAGTGGGTGGGAAATTAAGTAAATTTTGAACCATTCCTTAGTCATTTCCA 1685
Qy 1664 aacttgaacacagaatcagtttaaatctatttattctatttaagaaataactttgat 1723
Db 1686 AACTTGAACACAGATCAGTCTTAATATCATTTTATTTTATTAAGAAATTAATTTGAT 1745
Qy 1724 aaattaatcagccccagtgctttaaaaaattctcttaaataaaaaataatagactcgt 1783
Db 1746 AAATTAATCAGCCCGCAGAGTCTTTAAAAATTTCTTTAAATAAAATAATAGACTCGCT 1805
Qy 1784 agtcagtaagatatgatatgatatgctgccccctcagtgctcttgatcagatgac 1843

```

```

Db 1806 AGTCAGTAAAGATATTTGAATATGATATGTCGCCCCCTCCGGTGTCTTTGATCAGATGAC 1865
Qy 1844 atgtgcatttttcagaggagcgtgcagacagcgtggcattcttagattacttttttact 1903
Db 1866 ATGTGCGCATTTTTCAGAGGACGTGCAGACAGCGTGCATTTAGATTACTTTTCTTACTC 1925
Qy 1904 tgaacatgctgcttttggagtgaggattcaaaagtggtccaccgctgcccactactg 1963
Db 1926 TGAACATGGCGCTGTTTGGAGTGGCGGATTAAGAGTGTGCCACGCTGCCCTACTG 1985
Qy 1964 caatggcagttttaactcttcttttggcttctgcagatgggttgcaattgatccctaac 2023
Db 1986 CAAATGGCAGTTTAAATCTTTATCTTTGGCTCTGCAGATGGTGGCAATGATCTTTAAC 2045
Qy 2024 caataatgctcagtcctcactctgctcgtctcagtgccactgagtgccactgtgtttaag 2083
Db 2046 CAATAATGTCAGTCCCTCATCTCTGCTCTCTCATAGGTGCCACCTTGTGTGTTTAAAG 2105
Qy 2084 aagggaagctttgtaccttttagagtgtagtggaatgaatgaatggcttgagtgactg 2143
Db 2106 AAGGGAAGCTTTTACCTTTAGAGTGTAGTGAATGAATGAATGGCTTGGAGTGCAC 2165
Qy 2144 agaacagcatgattcttctgttggggaaagaaagaaatgctgctatgaaattgttggtg 2203
Db 2166 AGAACAGCATATGATTTCTTGTCTTGGGAAAGAAATGATGCTATGAAATGTGGGTG 2225
Qy 2204 gtgtattgagaataatcattgcttctgaatggagctgaattgataaaaccca 2263
Db 2226 GTGTATTTGAGAAGATAATCATTTGCTTATGTCATGTCATGTCATGTCATGTCATGTC 2285
Qy 2264 aaatacagctatgagtgctgggcaagtttacttttttctgtatgtttccctaac 2320
Db 2286 AAATCAGCATATGAGTGTGGGCAAGTTTACTTTTCTGATGTTCCTACAACT 2342

```

RESULT 12

```

LOCUS HUMHUGBR1 2351 bp mRNA linear PRI 07-MAR-1995
DEFINITION Human bilirubin UDP-glucuronosyltransferase isoform 1 mRNA,
complete cds.
ACCESSION M57899.1 GI:18472
VERSION M57899.1
KEYWORDS UDP-glucuronosyltransferase; bilirubin UDP-glucuronosyltransferase.
SOURCE Human adult female liver, cDNA to mRNA, clones 26, 211 and 26MB2.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2351)
AUTHORS Ritter, J. K., Crawford, J. M. and Owens, I. S.
TITLE Cloning of two human liver bilirubin UDP-glucuronosyltransferase
CDNAs with expression in COS-1 cells
JOURNAL J. Biol. Chem. 266 (2), 1043-1047 (1991)
MEDLINE 91093210
FEATURES
source
1..2351
/organism="Homo sapiens"
/isolate="AK"
/db_xref="taxon:9606"
/map="2"
/clone="26"
/sex="female"
/tissue_type="liver"
/dev_stage="adult"
/tissue_lib="lambda-ZAP"
1..2351
/gene="UGT1"
16..1617
/gene="UGT1"
/standard_name="bilirubin UDP-glucuronosyltransferase
isozyme 1"
/EC_number="2.4.1.17"
/codon_start=1
/product="UDP-glucuronosyltransferase 1"

```



```

Db 1901 GGCATTCTAGATTACTTTCTTACTCTGAAACATGCGCTGTTGGAGTGGGGATTCAA 1960
Qy 1938 agtggtccacgcgtgcccactactgcaaatggcagttttaattatttcttcttggctct 1997
Db 1961 AGGTGTCCTCCACCGCTGCCCTACTGCAATGCGAGTTTAACTTATCTTTTGGCTTCT 2020
Qy 1998 gagatggtgcaatgatccttaaccaataatggtcagctcctcctctctctctctc 2057
Db 2021 GCAGATGTTGCAATTGATCCCTTAACCAATATGGTCAGTCTCATCTGCTCTCTTC 2080
Qy 2058 ataggtgcaacctgtgtgtttaaaggaggagccttbttaaccttagagtgtagtgaa 2117
Db 2081 ATAGGTGCGACCTGTGTGTTTAAAGAGGCAAGCTTTGTACCTTTAGAGTGTAGTGAA 2140
Qy 2118 atgaatgaatgcttgagtgactgagacagacagcatatgttctgtcttggggaata 2177
Db 2141 ATGANTGAATGCTTGGAGTGACAGACAGCATATGATTTCTGCTTGGGGAATA 2200
Qy 2178 gaatgatctcgaatggtggtggtggtgtatttgagaagataatcatctcttatgcaa 2237
Db 2201 GAATGATGCTATGAATTTGGTGGGTGTTATTTTGAAGAATAATCATCTGTTATGCAA 2260
Qy 2238 atggagctgaattgtataaaacccaaatacacagctatgaagtcgtggcaagtttact 2297
Db 2261 ATGGAGCTGAATTTGATAAAACCCAAATACAGCTATGAAGTGTGCTGGCAAGTTTACTT 2320
Qy 2298 ttttctgatgttctctacaact 2320
Db 2321 TTTTCTGATGTTCTCTACAACCT 2343

RESULT 11
BC019861 2388 bp mRNA linear PRI 22-JAN-2002
LOCUS Homo sapiens, clone MGC:29860 IMAGE:4454322, mRNA, complete cds.
DEFINITION BC019861
ACCESSION BC019861
VERSION BC019861.1 GI:18043700
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2388)
Strausberg, R.
Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Guanaraine, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 42 Row: c Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4507814.
location/Qualifiers
1. .2388
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="MGC:29860 IMAGE:4454322"
/tissue_type="kidney, hypernephroma"
/clone_lib="NIH_MGC_89"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
35..1633
/codon_start=1
/product="unknown (protein for MGC:29860)"
/protein_id="AAH19861.1"
/db_xref="GI:18043701"
/translation="MACLLRSFORISAGVFFLALMGWVVDKLLVVPQDGGSHLMSMD
IVEVLSDRGHEIVVVPVNNLLKESKYTRKIYVPYDOELKNRYOSFGNNHFAER
SEFUTAPOTYENNNMIVIGLYFINCSLLQDRDITLNFKEKSDALFTDPALPCCVILA
EYLGLPSVLYFRGFPSCSLEHTSPDPVSIPRCYTKSDHMTFSORVANFLVNLLE
PYLFCLESKYVELASAVLRDVIITLYQKVSVLLRVDVLEVPRVPMNVMVFIGG
INCKRKDLISQEFFEAYINAGSEHIVFESLGSMSVSEIPEK KAMATADALGKIQOTVLM
RYTGFRPNLANNTILYKWLQNDLLGHPTWRAFIHAGSHGVYESICNGVPMVMPL
FGDDMDNAKMETKAGVTILNLEMTSEDLNALKAVINDKSTYKENIMRSLSHKDRP
VEPDLIAVFWVEFYVNRHKGAPHLRPAAHDLTWQYHSLDVIGFLAVVLTVAFTFKC
CAYGYRCLGKGGKRVKKAHKSHT"
BASE COUNT 648 a 503 c 546 g 691 t
ORIGIN

Query Match 70.2%; Score 1629; DB 9; Length 2388;
Best Local Similarity 82.5%; Pred. No. 0;
Matches 1879; Conservative 0; Mismatches 395; Indels 3; Gaps 1;

Qy 47 ttctctatgtgtgtctgtcgtgacctgtgcttgcgagtgagcagggaagctactgg 106
Db 66 TTTCTGCACGGGTTTCTTCTTAGCACCTTTGGGCGCATGTTCTAGCTGACAAGCTGCTGG 125
Qy 107 tagtcccatgatgagcactggtccaccatgagctgagtggtgagaaactcatic 166
Db 126 TGTCCCTCAGGACGAGCGACTGCTGTAGTAGAAGGATATAGTTGAGGTCTCAGTG 185
Qy 167 tcagggggcatgagtggtgttagtcatgcccagagtgagttggcaactgggaagatcac 226
Db 186 ACCGGGGTCATGAGATTGTAGTGGTGGTGGCTGAAGTTAATTTGCTTTGAAAGAAATCCA 245
Qy 227 tgaattgcacagtgagacttattcaactcatatccctgagagatctgagccggagt 286
Db 246 AATACTACAGAAAGAAATCTATCCAGTGGCGGTATGCCAAGAAAGAGCTGAAGACCGTT 305
Qy 287 tcaaggtctttggccatgctcctaaaggagagcttctttgatgacagtggttctcgate 346
Db 306 ACCAATCATTTGGAAACAATCTTTGCTGAGCGATCATTCCTAACTGCTCTCAGACAG 365
Qy 347 tgggttcatacaatgaca---ttttgacttatttttttcaaatgagaggtttgttta 403
Db 366 AGTACAGGATTAACATGATTGTTATTGGCCTGTACTTCACTCACTGCCAGAGCTCTCTGC 425
Qy 404 agacacacacacacacacacacacacacacacacacacacacacacacacacacacac 463
Db 426 AGSACAGGACACACCTCAACTCTTTTAAAGGAGAGCAAGTTTGTATGCTCTTTTACAGACC 485
Qy 464 cttttgataactgtggcttaattgttgccaaatatttctccctccctccgtggtcttgg 523
Db 486 CAGCCTTACCTGTGGGGTGATCCTGGCTGAGTATTTGGGCTTACCATCTGTGTACTCT 545
Qy 524 ccaggggatcatttgcacatacttgagaagggtgcacagtgccctgctctctctctct 583
Db 546 TCAGGGGTTTTCGTGTTCCCTGGAGCATACATTTACAGAGAGAGCCAGACCTGTGCTCT 605
Qy 584 atgtcccccgaattctcttagggttctcagatgccatcttctcagagagagatcgga 643
Db 606 ACATTTCCCGAGTGCTACACAAAGTTTTCAGACCACATGACTTTTCCCAACAGAGTGGCCA 665
Qy 644 accacatctgacattgaggaggaacatttattatgccaccgttttttcaaaaatgcctag 703
Db 666 ACTCTCTTTTAAATTTTGGAGCCCTATCTATTTTATTTTGTCTGTCTTCAAGATGAAG 725
Qy 704 aaatgacctctgaaattctccaaacacacctgttaccgagtgatgatctctacagcacat 763

```

LAKYLSPVAFVFWRYIPCDLDFKQCPNPSSVYIPKLLTNSDHTFLQVRKNLYPL
 ALSYICHTFSAFYASLASLSEFOREVSYDLVSYASVWLFGRDFYMDYPRPIMPNAVFI
 GGINCANGKPLSQEFAYINASGEHGVVFSLSGMSVSEIPEKKKAYADALGKIPQTV
 LMRVYTRPNSLANNTILVKMPLQNDLLGHPMTAPITFAGSHGVYESICNGVPWVM
 PLFGDMDNARMETKAGVTLNMLEMTSBDLENLAKAVINDKSYKENIMRLSLHLD
 RVEPEPLDAVFWVEFYMRHKGAPHLRPAADHLTWQYHSLDVIQGLLAVLVIAFTF
 KCAVGVKRLGKGLKGRVKKAKSKTH"

sig_peptide

/gene="UG2B"

/note="G00-127-753"

96..1631

/gene="UG2B"

/BC_number="2.4.1.17"

/note="G00-127-753"

/product="UDP-glucuronosyltransferase 2 family polypeptide
 B"

BASE COUNT 504 a 534 c 573 g 657 t
 ORIGIN

Query Match 70.5%; Score 1636.6; DB 9; Length 2368;
 Best Local Similarity 83.4%; Pred. No. 0;
 Matches 1871; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

QY 81 ctttgcagaggaaggaactgtagtgcacatgagtgagagcaactgtttcacocat 140
 DB 101 CTGGGCTGAGAGTGGAAAGGTGTGGTGGTGGCCCATGATGGCAGCCCTGGCTCAGCAT 160
 QY 141 gaggtcggtgtgtgagaaactcatctcagggggcagatgagtggtgtgtgtagtcatccaga 200
 DB 161 GCGGAGGCCCTTGGGAGCTCCATGCCAGNGGCCACCAGCGGTGGTCTCACCCGAGA 220
 QY 201 ggtgagtggaactgggaagatcaactgaattgcacagtggaagactattcaacttcata 260
 DB 221 GGTGAATATCATCAATAAGAGAGAAATTTTTCACCTGTGACAGCTATGCTGTTCATG 280
 QY 261 taccctggagatctggaccgggaggttcaaggcttttggccatgctcaat---ggaagc 317
 DB 281 GACCCAGAGAAATTTGATCGCTTACGCTGGCTACACTCAGGGTCTTTTGAACACAGA 340
 QY 318 acaagtcagaagtatatattctctatttaattgggttcatacaatgacattttgacttatt 377
 DB 341 ACATCTTCTCAGAGATATTTAGAGTATGGCAATTTAACAATATGATCTATCTTTGGCCCT 400
 QY 378 tttttcaaatgtcaggagttgttttaagacaaataatttagatagatacttaaaagagag 437
 DB 401 TCATAGGTGTGTGGAGCTACTGCATAATGNGCCCTGATCAGGCACCTGAATGCTAC 460
 QY 438 ttcttttgatgcagtggtttctcgatcttttgataactgtgtggttaattgttgcacaata 497
 DB 461 TTCCTTTTGTGTGTGTAAACAGACCCGCTTAACCTCTGCGGGCGGTGCTGGCTTAAGTA 520
 QY 498 tttctccctccctccgtgtgttcgcaggggaaatactttgcaactatcttgaagaagg 557
 DB 521 CCTGCTGATTCCTGCTGTGTTTTTGGAGGTACATTCATCCATGTGACTTAGACTTTAAGG 580
 QY 558 tgcacagtgccctgtctcttctctatgttcccccagaattcttcttgggtttctcagatgc 617
 DB 581 CACACAGTGTGCAAAATCTTCTCTCTATATTTCTTAAGTACTAACGACCAATTCAGACCA 640
 QY 618 catgactttcaagagagagtaaggaaccacatcatgcaacttggaggaacatttattatg 677
 DB 641 CATGACATTCCTGCAAGGGTCAAGACATGCTCTACCCCTGCGCCCTGTCTCATATTG 700
 QY 678 ccacggttttttcaaaaatgccctagaaaatagcctctgaattcttccaaacacctgttac 737
 DB 701 CCATACATTTTCTGCCCCCTTATGCAAGTCTTCGCTCTGAGCTTTTTCAGAGAGAGGTGTC 760
 QY 738 ggagtatgactctacagccacacataatttgggttggcgaacggaactttgttttggga 797
 DB 761 AGTGGTGAATCTGTGACGTATGATCCGTTGGCTGTTCGAGGGGACTTTGTGATGGA 820
 QY 798 ctatcccaaacccgtgtagtgcacacatgatcttcttcttgggttatcaactgcatcagg 857

DB 821 CTACCCAGCGCGGATCATGCCAACATGTGCTCTCATTTGGGGCATCAACTGTGCCAACGG 880
 QY 858 aaagcgttgccataggaatttaagacctacatttaactctcttgagaaacatggaattgt 917
 DB 881 GAACCCACTATCTCAGGAATTTGAAGCCTACATTAATGCTTCTTGGGAACATGGAATGT 940
 QY 918 ggttttctcttgggatcaatggtctcagaataatccagagaagaagactatggcaattgc 977
 DB 941 GGTTTTCTCTTTGGGATCAATGTGCTCAGAAATTCAGAGAAGAAAGCTATGCGCAATTGC 1000
 QY 978 tgaagcttgggcaaaatccctcagacagtcctgtgcsgtacacttggaaccccaacatc 1037
 DB 1001 TGATGCTTTGGGCAAAATCCCTCAGACAGTCTCTTGGCGGTACACTGGAAACCCGACCATC 1060
 QY 1038 gaattcttggaacaacacagatgacttgaagtggctaccccaaaacagatctgttggta 1097
 DB 1061 GAATCTTGGCAACACAGATACTTGTAAAGTGGCTACGCCCAAAACGATCTGCTTGGTCA 1120
 QY 1098 ccgagtgaccgtgctttatcacccatgctgttcccatggttccatggtttalgaagcatatg 1157
 DB 1121 CCCGATGACCCGTGCTTTATCACCAATGCTGTGCTTCCCATGTGTTTATGAAGCATATG 1180
 QY 1158 caatggcgttcccatggtgtgatgcccctgttgggtgatcagatggacaatgcgaagcg 1217
 DB 1181 CAATGGCGTTCCTCATGTGATGATGCCCTGTTTGTGATCAGATGGACAATGCAAGCG 1240
 QY 1218 catggaactaaaggagctggagtgaccctgaatctctgaaatgacttctgaagattt 1277
 DB 1241 CATGGAGACTAAGGAGCTGGAGTGCACCTGAATGTTCTGGAATGACTCTGGAAGATTT 1300
 QY 1278 agaaaaatgctctaaagcagtcacatgacaaaagtacaaaggagagaacatcagcct 1337
 DB 1301 AGAAATGTCTTAAAGCAGTCAATCAATGACAAAAGTTTACAAGGAGAATCATCGCGCT 1360
 QY 1338 ctccagcttccaaagagaccccggtgagcgcgtgagcgcgtgacactgcccgtctctgggtgga 1397
 DB 1361 CTCAGCCTTCACAGAGACCGCGGTGGAGCGCTGGACCTGGCCGTGTTCTGGGTGGA 1420
 QY 1398 gtttgtgatggagcaaaagggcgccacactgcccccgagccacagcactcaactg 1457
 DB 1421 GTTTGTGATGAGGCAACAAGGCGCGCACCTGCGCGCGCGCAGCCAGACCTCACCTG 1480
 QY 1458 gtaccagtcacactctcttgagcgtgattggttctctcttggcgtgctgtcagagggc 1517
 DB 1481 GTACCAGTACCAATCTTGGACGTGATTGGTTTCTCTTGGCCGTGCTGCTGACAGTGGC 1540
 QY 1518 ctteatcaactttaaattgtgtctatggtaccgggaatgcttggggaaaaaaggcg 1577
 DB 1541 CTTTCATCCTTTAAATTTGTGCTTATGGCTACCGGAAATGCTTGGGAAAAAAGGCG 1600
 QY 1578 agttaagaaagcccaaaatccaagaccattgagaagtgggtggaataaggtaaat 1637
 DB 1601 AGTTAAGAAAGCCCAAAATCCAAGACCCATTTGAGAGTGGTGGGAAATAGATTAAT 1660
 QY 1638 ttgtaaccattcccttagtctatttccaaacttgaacacagaatcagttgttaattcatttt 1697
 DB 1661 TTTTGAACCATTTCCCTAGTCAATTTCCAAACTTGAAACAGAAATCAGTGTAAATTCATTT 1720
 QY 1698 attctattaaagaaatactttgcataaataatcagccccagagtgctttaaaaaatc 1757
 DB 1721 ATTCCTATTAAAGAAATCTTGCATAAATTAATCAGCCCCCAGAGTGTCTTAAATTCATTT 1780
 QY 1758 tcttaataaaaaataatagactcgttagtcagtaagaatatattgaatatgtatgctgcc 1817
 DB 1781 TCTTAATAAAAAATAATAGACTCGCTAGTACAGTAAGATATTTGAATATGATCGTGCCC 1840
 QY 1818 cctccggttctttagatcaagatgacatgtgccatttttcagagagcgtgagacagcgt 1877
 DB 1841 CTTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGTGACAGAGCT 1900
 QY 1878 ggcattcagatgacttttcttactctgaaacatggcctgtttggagtgctgagattcaa 1937

Db 265 ATTCAACTTCATATA - CTTGAGGAGATCGACCGGAGTTCAGAGGCTTTTCCCATGCTC 323
 QY 308 aatgaaagcaagaagtaagagatataatctctataataatgggttcaataatgacattt 367
 Db 324 AATGGAAGACCAAGTAGAAGATATATTCCTATTAATGGGTTCATACAAATGACATTT 393
 QY 368 ttgacttatttttttcaaatgagaggtttgttttaagacacaaaaaattagtagaataact 427
 Db 384 TTGACTTATTTTTTCAAAATGACAGGAGTTTGTCTTAAAGACAAAAAATAGTAGAATACT 443
 QY 428 taaagagaggtctttttgatgacagtggtttctcgatccttttgataactgtggttaattg 487
 Db 444 TAAAGGAGAGTCTTTTTGATGACAGTGTCTCGATCCCTTTTGATATACTGTGCTTAATG 503
 QY 488 ttgcaaaatatttccctccctccctccgtgttcttgcagaggaataactttgccaactatc 547
 Db 504 TTGCCAAATATTTCTCCCTCCCTCCGTGTCTTGCACGAGGAATACCTTTGCCACATATC 563
 QY 548 ttgaagaaggtgcaagtgccctgctcctcttctctatgtcccccagaattctcttaggt 607
 Db 564 TTGAAGAAGGTGCACAGTGCCTGCTCTCTTCTATGTGTCCTCCAGAAATCTCTTAGGGT 623
 QY 608 tctcagatgcacttctcaagagagagtagtacggaacacacatcatgctacttggagaac 667
 Db 624 TCTCAGATGACATGACTTTCAAGGAGAGATACGGAAACACATCATGCTTTGGAGGAAC 683
 QY 668 atttattatgcacagtttttcaaaaatgccctagaaatagcctctgaaattctccaaa 727
 Db 684 ATTTATTTATGCCACCGTTTTTCCMAAATGCCCCCTAGAAATAGCCTCTGAAATTCCTCANA 743
 QY 728 cactgttgaagagtatgatctctacagccacacatcaatttgggtgttgcgaacgact 787
 Db 744 CACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTCGAGAGGACT 803
 QY 788 ttgttttggaactcccaacccgtgatgccaacatgatctctcaatttgggtatcaact 847
 Db 804 TTGTTTGGACTATCCCAACCCGTGATGCCCAACATGATCTTCAATGTTGGTATCAACT 863
 QY 848 gccatcagggaagccgttgcctatggaatttgaagcctacattaatgotttggagaac 907
 Db 864 GCCATGAGAGGAAGCGTGTGCTATGGAAATTTGAAGCCTACATTAATGCTCTTGAGAAC 923
 QY 908 atggaatttgggttcttcttggatccaatggtctcgaataatccagagaagaagcta 967
 Db 924 ATGGAATTTGGGGTTCTCTTTGGGATCAATGCTCTCAAGAAATTCAGAGAAAGAAAGCTA 983
 QY 968 tggcaattgtgatgctttgggcaaaatccctcagaagctctgttggcggtacactgaa 1027
 Db 984 TGGCAATTTGTGATGCTTTGGGCAAAATCCCTCAGACAGTCTCTGTGGCGGTACACTGGAA 1043
 QY 1028 cccgaccatcgaatcttgcgaacacacagatacttgaatggctaccaccaaaacgac 1087
 Db 1044 CCGGACCATCAATCTTCGGAACAAACACGATACCTTTGTTAAGTGGCTACCCAAAACGATC 1103
 QY 1088 tgcctgttaccagatgaccctgtcctttatacccaatgctgggttcccaatggtttatg 1147
 Db 1104 TGCTTGGTACCCGATGACAGTGCCTTTTATCACCATGCTGGTTCCTCATGTTGTTATG 1163
 QY 1148 aaagcataatgcaatggcgttcccatggtgatgacgacctgttgggtgatcagatggaca 1207
 Db 1164 AAAGCATATGCAATGGCGTTCCCATGTTGATGATGCCCTTGTGTTGGTATCATGAGTGGACA 1223
 QY 1208 atgcaaacgcatggaactaaggaactggaactggaactggaactggaactggaactggaact 1267
 Db 1224 ATGCNAAGCGCATGGAGACTAAGGAGCTGGAGTGCACCATGAATGTTCTGGAAATGACTT 1283
 QY 1268 ctgaagaatttagaaaatgctctaaagcagctcatcaatgacaaaaagttaacagggagaaca 1327
 Db 1284 CTGAAGAATTAGAAAAATGCTCTAAAGACAGTATCAATGACAAAAAGTTACAGGAGAAC 1343
 QY 1328 tcattgcgctctcagccttcaagagacccggtgagccgctgagccgctgagccgctgagccgct 1387

Db 1344 TCATGCGCCTCTCCAGCCTTCAACAGGACCCGCCGTTGGAGCGCTGGACCTGGCCCTGT 1403
 QY 1388 tctgggtgaggttctgtgatgagccacaaaggcgccacacotggcgcccccagccaccac 1447
 Db 1404 TCTGGGTGAGTTGTTATGAGGCACNAGGGCGGCACACACCTTGCGCCCGCACG 1463
 QY 1448 acctacctggtacacagtcacattccttggacggtgattgtttcctcttggcctgtgac 1507
 Db 1464 ACCTACCTTGGTACCAGTACCATCTTGGAGCGTATGTTGCTTTTCCCTTGGCCCGCTGC 1523
 QY 1508 tgacgtggcctcatcaccttttaaatgttgtgttattgcttaccggaatgcttgggga 1567
 Db 1524 TGACAGTGGCTTCATCACCTGTAATGTGTGCTTATGCTTACCGGAAATGCTTGGGGA 1583
 QY 1568 aaaaagggcgagtttaagaaagcccaaatccaaagacccattgagaagtgggtgggaaat 1627
 Db 1584 AAAAAGGGCGAGTTAAGAAAGCCCCCAAAATCCAAGACCCCATTTGAGAGGTGGGGGAAAT 1643
 QY 1628 aaggtaaaattttgaaacattcccttagtcatttccaaacttgaacacagaatcagtgta 1687
 Db 1644 AAGSTAAATTTTGAACCA - TCCCTAGTCAATTTCCAAACTTTGAAAAACAGATCACTGTTA 1702
 QY 1688 aattcaatttattcttattaaagaaataactttgcaataataataatcagccacagtg - ct 1746
 Db 1703 AATTCAATTTTATCTTATTAAGGAAATACCTTTGCAATAAATTAATACGCCCGAGGTCT 1762
 QY 1747 ttaaaaaattctcttaaaaaataatagactcgtcgtcagtaagaatatttgaatat 1806
 Db 1763 TTAATAAATTTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATAT 1822
 QY 1807 gtatcgtgccccctccgtgtctttgatcagatgacatgtgccatttttcagagggcgt 1866
 Db 1823 GTATGTTGCCCCCTCCGGAGTCTTTTGTATCAGATGACATGTGCCATTTTCAAGGACGT 1882
 QY 1867 gcagacaggtcggcattc - tagattactttctactctgaaacatggcctgtttggag 1925
 Db 1883 CGACAGGCTGGCATTTCTTAGATTACTTTCTTACTCTGAAACATGGCTGTTTGGGAG 1942
 QY 1926 tgcgggatccaagtggtccaccctgcccctactcactcgaatggcagttttaaacttat 1985
 Db 1943 TCGGGATTCAAAGGTGTCCACGCTGCCCTACTGCAAAATGGGAGTTTAAATCCCTAT 2002
 QY 1986 ctttggctctgcagatggttgcattgatccttaacccaaataatggcagccctcatct 2045
 Db 2003 CTTTGGCTCTCGAGATGGTTCG - ATTGATCCTTAACCAATAATGTTGTTCTTCTATCT 2061
 QY 2046 ctgctccttcataggtgcccacctgtgtttaaagaagggaagcctttgtacacttag 2105
 Db 2062 CTGCTCTGACTTCATAGGTGCGAGGTGTGTTTAAAGAGGGAAAGCTTTGTACCTTTAG 2121
 QY 2106 agttaggtgaaatgaatgaatggttggcttggagtgcaactgagaacagcatatgatcttgc 2165
 Db 2122 AGTGTAGGTCAATGAATGAATGGCTTGGAGTGCACCTGAGAACAGCATATGATTTCTTGC 2181
 QY 2166 ttggggaaaaaagaatgatgctatgaaattgggtggtgtattttgagaagaataatcat 2225
 Db 2182 TTTGGGAAAAAGCATGATGATGAAATTTGGTGGTGTGTTTATTGAGAAGATAATCAT 2241
 QY 2226 tgcctatgcaaatggagctgaatttgataaaaaaaccccaatacagctatgaagtgcgg 2285
 Db 2242 TGCTTATGTCAAATGGAGGTTAATTTGATAAAAACCCAAAATACAGTATGAAGTGTGG 2301
 QY 2286 gcaagtttac - ttttttctctgatgtttccctacaact 2320
 Db 2302 GCAAGTTTACTTTTTTCTGATCTTTTCTCTACAAAT 2337

RESULT 8

AF104336

LOCUS

DEFINITION Macaca fascicularis UDP-glucuronosyltransferase UGT1A09 mRNA,
 complete cds.

ACCESSION AF104336

2396 bp

mRNA

linear

PRI 13-DEC-1999

QY 248 a

248 at

Query Match	93.0%	Score 2156.8	DB 9	Length 2392	
Best Local Similarity	97.7%	Prod. No. 0			
Matches 2262	Conservative	0	Mismatches 47	Indels 7	Gaps 7
9	cagttctctatggtcttcacagggtagcagccctctctctatgtgtgtctgct	68			
25	CACCTCTGGCATGGCTTGCACAGSGTGACACAGCCCTTCTCTATGTGTGTGTCTGCT	84			
69	gctgaactgttggtctttccggagggcagggaagctactgttagtgcccatgagtgagccca	128			
85	GCTGACCTGTGGCTTTGGCAGGCGAGGGAAGCTAGTGGTAGTGCCCATGGATGGCAGGCCA	144			
129	ctygttcacatgaggtcgggtggttgagaaactcatcttcaggggcgcatcagggtggtgtgt	188			
145	CTGTTTCACATCAGGTGGTGGTGGAGAAACTCATTTCTCAGGGGGCATCAGGTGGTGT	204			
189	agtcac-tggccagaggtgagttgycgaactgaggagaatacactgaattgcacagtgagacatt	247			
205	AGTCAGTGCCAGAGGTGAGTTGGCAACTGGGAAGATCAATGAATTGCACAGTGAAGACTT	264			
248	attcaacttcataccctcagaggtactctgaccgggaggttcaaggcgtttttgcccactgc	307			

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens, Similar to UDP glycosyltransferase 1 family,
polypeptide A9, clone MGC:9353 IMAGE:3850654, mRNA, complete cds.
BC020971
BC020971.1 GI:18088086
MGC.
human.

REFERENCE
AUTHORS
TITLE
JOURNAL

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2423)
Strausberg, R.
Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 20 Row: m Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomeScan gene
prediction, similarity but not identity to protein.

FEATURES
source

Location/Qualifiers
1..2423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:9353 IMAGE:3850654"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
70..1662
/codon_start=1
/product="Similar to UDP glycosyltransferase 1 family,
polypeptide A9"
/protein_id="AAH20971.1"
/db_xref="GI:18088087"
/translation="MARAGWTSVPVLCVLLTCTGFAFGKLLVVPMDGSHWFTMQSV
VEKILRGHEVYVYVMEYSNQLRSNCTVKTYSYSLTDQNRFPVFAHAKWAKQA
QSIPLMSSSGFLDLPFHSRSLFNDRKLYELKESDFDVLDPDFDTGLIVAKY
FSLPVVYTRGIPCHLEEGACQAPLSPVNDLIGFSDAMTKERVMNHIHLEHDL
FCOGLPFEAFYINASCEHGIWVFSIGMSVSEIPEKKAWIADALGKIPOTVWRY
TGTRPSNLANTILYKPLPNDLLGHPTAPFTHAGSHGVESICNGYPMVMPLEPC
DMDNARHETGAGTVLNLVEMTSEDLNALKAVINDKSTKENIMRSLSLKHORPVE
PDLIAVWFVFWRHGAPHLRPAADHLTWYQYHSLDVIGFLAVLVITVAFITFCCA
YGRKCIIGKKGRVKAHKSHTH"

CDS

BASE COUNT 643 a 514 c 559 g 707 t
ORIGIN

Query Match
Best Local Similarity
Matches 224/

94.68; Score 2194.6; DB 9; Length 2423;
97.08; Pred. No. 0;
0; Mismatches 69; Indels 1; Gaps 1;

QY

5 gctgcagttctctgtgacaggttgaccagccccctctctctctctgtgtgtc 64

Db

56 GCTGCAGTCTCTCATGCTCGCGAGGGTGGACGAGCCCGCGTCTCTTATGTGTGTGTC 115

QY

65 tgetgtacctgtggtcttgcagagggaggaagctactgtgtgtgccccatgatgga 124

Db 116 TACTGCTGACCTGTGGCTTTCGAGGAGGAGAGCTGCTGTAGTCCCATGGGGA 175
QY 125 gccactggttcacacatgagtcggtggtgagaaaactctctcagggggcagtgaggtgg 184
Db 176 GTCACCTGGTTTCCACCATGCAGTCGGTGGTGGAGAACTTATCTCAGGGGGCATGAGTGG 235
QY 185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaga 244
Db 236 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGNAAGATCAGTGAATTCACAGTGAAGA 295
QY 245 ctattcaacttcataacccgtggaggtctggaccgggaggttcgaaggtcttggccatg 304
Db 296 CTTACTCAACCTGTCACACTCTGGAAGATCAGAACCGGGAATTCATGTTTCCGCCATG 355
QY 305 ctcaatgaaagacaagtcaggaagtatatctctatttaaatgggttcatacaatgaca 364
Db 356 CTAATGGAAGACACAGGCACAAAGTATATTTCTCTATTAATGAGTTCATCCAGTGGTT 415
QY 365 ttttgacttatttttcaaatgcagggtttgtttaaagacaaaaaattagtagaat 424
Db 416 TTCTTGACTTATTTTTCGCATTGTCAGGAGTTGTTTAAATGACCGAAAATTAGTAGAAT 475
QY 425 acttaaaagagagttttttgatgcaggttttctcaatctcttcttaactaactgtgcttaa 484
Db 476 ACTTAAAGAGAGTCTTTTGTATGTCAGTCTTCTGATCTCTTCTTATGACCTGCTTAA 535
QY 485 ttgtgcacaaatattctccctccctccctggtgtcttcgccaggggataacttgcacct 544
Db 536 TTCTTGTCTAAATATTTCTCCCTCTGCTGCTTTCACAGGGGGAATATTTTGGCACC 595
QY 545 atctgaagagagtcacagtcacctgctctctctctctctctctctctctctctctctctag 604
Db 596 ATCTTGAAGAGGTGACAGTGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAG 655
QY 605 ggtctcagatgacctgaacttcaagagagagtaagtaagcaacacatcatgcaacttgag 664
Db 656 GGTTCACAGATGCCATGACTTTTCAAGAGAGAGATATGGAACACACATCGTCGCTGGAG 715
QY 665 aacatttattatgcacccgttttttcaaaaatgcctagaataagctctctctctctctctc 724
Db 716 ACATTTATTTGGCAGTATCTTTTGTAGAAATGCCCTAGAAATAGCTCTCAAAATCTCC 775
QY 725 aaacacctgttacggagtagatgactctacagccacacatcaatttgggtgtgtggaacgg 784
Db 776 AAACCCGTCTCACGGCATATGATCTCTACAGTCACACATCAATTTGGTGTGGGAACGG 835
QY 785 acttgttttggactatcccaaacccgtgatgcccaacatgactctctctctctctctctc 844
Db 836 ACTTTGTTTGGACTATCCCAACCCGTGATGCCCAACATGATCTTCATTTGGTGTATCA 895
QY 845 actgcatcaggggaagcccttgcctatggaattgaagcctacataactgtctctgag 904
Db 896 ACTGTCAACGGAAGAGCCATTCCTTATGGAATTTGAAGCCTACATTAATGCTTCTCTGGAG 955
QY 905 aacatgaaattgtgtgtttctctcttggatcaatggtctcagaaattccagagagaagaag 964
Db 956 AACATGGAATGTGTTTCTCTTTTGGGATCAATGGTCTCAGAAATTCAGAGAGAAG 1015
QY 965 ctatggcaattgtgtgtcttgggcaaaatccctcagacagtcctgtgtggtgtgtgtgtg 1024
Db 1016 CTATGGCAATTCGTGCTCTTTGGGCAAAATCCCTCAGACAGTCCCTGCGGGGTACACTG 1075
QY 1025 gaaccgcacatcaactcttcgcaacaacacagatgacttgttaagtggctacccccaaacg 1084
Db 1076 GAACCCGACCATCGAATCTTTCGCAACACACAGATACCTGTTTAAAGTGGTGTACCCAAAACG 1135
QY 1085 atctgtgtgtcaccacagatgacccgtgctcttattcaacacatgctgtgtgtgtgtgt 1144
Db 1136 ATCTGTGTGTCACCCGATGACCCGTGCTTATCACCACCATGCTGTGTGTGTGTGTGTGT 1195
QY 1145 atgaagcatgataagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1204

Db 230 TTCTAGTCATGCCAGAGGTGAGTGTGGCAACTGGGAAAAATCACTGAATTCACAGTGAAGA 289
 QY 245 ctattcaacttcataacacctggaggtatctggaccggaggttcaaggcttttgcccaatg 304
 Db 290 CTTTACTCAACCTCATACACTCTGGAGGATCTGGACCGGGAATTCATGATTTCCGCGNTG 349
 QY 305 ctcaatggaagacacaggtacagatataatctctattatctctattatgggttcatacaagaca 364
 Db 350 CTCATGGAAGACACAGTACAGAGTTGTGTTTCTCTATTTCTCTGAGTTCATCAATGAGTT 409
 QY 365 tttttgaacttatttttcaaatgacagagttgttttaaaagacaaaaaattagtagaat 424
 Db 410 TTTTAACTTATTTTTCGCATTTGCAGGAGTTGTGTTTAAATGACCGAAAAATTAGTAGAAT 469
 QY 425 acttaaaagagagttcttttgatcagtggttttcgatctcttttgaataaacttggtcttaa 484
 Db 470 ACTTAAGGAGAGTCTTTTGATGCGGTTTCTTGATCTTTTGATGCCGTGGCTTAA 529
 QY 485 ttgttgcaaaatatttccctccctccctggtgcttcgccggggaatacttggcaat 544
 Db 530 TTGTTGCCAAATATTTCCCTCCCTCTGTGTGCTTTCGCCAGGGGAATAGCTTGGCACT 589
 QY 545 atcttgaagaaggtgacagtgccctgctctcttcttctctatgctcccgagaattctctag 604
 Db 590 ATCTTGAAGAAGTGCACAGTGCCTGCTCTCTTCTTCTTATGTCCCCAGAAATCTCTTAG 649
 QY 605 ggttctcagatgcacatgacttcaaggagagagtacggaaacacacatcatgcacttggag 664
 Db 650 GGTCTCAGATGCCATGACTTCAAGAGGAGAGTACGGAACACATCATGCACTTGGAGG 709
 QY 665 aaacattattatgcccacgctttttcaaaaatgcccctagaaaatagcccttgaattctcc 724
 Db 710 AACATTTATTTTGGCAGTATTTTCCAAAAATGCCCTAGAAATAGCCCTGAAATTCCTCC 769
 QY 725 aaacacctgtacagagatgatctctacagcaacacatcaaatgtgtgttgcgaag 784
 Db 770 AAACACCTGTCCAGCATGATGATCTACAGCCACACATCAATTTGGTGTGCGAAGC 829
 QY 785 acttgttttggactcccaaacccgtgatgcccaacatgactctcattgggtgatca 844
 Db 830 ACTTTGTTTGGACTATCCCAACCCGCTGATGCGCCCAATATGATCTTCATTTGGTGTATCA 889
 QY 845 actgcatcagaggaagcgtgtgcttatggaattgaaacctcaaatgaatgcttctgag 904
 Db 890 ACTGCCATCAGGGAAGCCATTTGCCATGAGNAATTTGAACCTTACATTAATGCTTCTGGAG 949
 QY 905 acatggaattgtggtttctctcttggggtcaatggtctcagaattccagagaagaag 964
 Db 950 AACATGGAATTTGTGTTTCTCTTTGGGATCAATGGTCTCGAATAATTCAGAGAAGAAAG 1009
 QY 965 ctatggcaattgtgatgcttttgggcaaaatccctcagacagtcctgtgcggtacatg 1024
 Db 1010 CTATGCAATTTGCTGATGCTTTTGGCAAAATCCCTCAGACAGTCTCTGTGGCGGTACACTG 1069
 QY 1025 gaacccgacatcgaatctcttgcaaacacagatacttggtaagtggtctaccccaaacg 1084
 Db 1070 GAACCCGACCATCGAATCTTTGGAACACACAGATACTTTTAAAGTGGCTACCCCAAAAG 1129
 QY 1085 atctgtcttgggtcaccgcatgacccgtgctttatcaccatgctggttcccatggtgttt 1144
 Db 1130 ATCTGCTTGTGTCACCGGATGACCCGTGCCCTTATACCCATGCTGGTGTCCCATGCTGTTT 1189
 QY 1145 atgaagcatatgcaatggcgttcccatgggtgatgagatgaccccttgggttggatcagatgg 1204
 Db 1190 ATGAAGCATATGCAATGGCGTTCCCATGTTGATGATGTCCTTGTGTTGGTGTATCAGATGG 1249
 QY 1205 acaatgcaagcgcacatggagactaaggagctgaggtgagtgacccctgaatgttctggaatga 1264
 Db 1250 ACAATGCAAGCGCATGGAGACTAAGGAGCTGGAGTACCCCTGAATGTTCTGCAAAATGA 1309
 QY 1265 cttctgaagatttagaaaaatgctctaaagcagtcacatgacaaaagttaacagagaga 1324
 Db 1310 CTTTCTGAAGATTTAGAAAAATGCTCTTAAAGACAGTCAATCAATGACAAAAAGTTACAAGGAGA 1369

QY 1325 acatcatcgctctccagccttcacaagagaccgcggtgagcgctggaactgacatgacg 1384
 Db 1370 ACATCATGCGCTCTCCGCTTCACAGGACCCGCCGTGGAGCGCTGACCTGGCCG 1429
 QY 1385 tgtctggtgtgaggtttgtgtgaggtgcaagggtgcgcacacactgccccgcagccc 1444
 Db 1430 TGTTCGTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCACACCTGGCCGCCAGCC 1489
 QY 1445 aqacacctccctgggtacacagttaccatctcttggagctgagtgtgttctcttggccgtg 1504
 Db 1490 ACGACCTCACCTGGTACCAGTACCAATTCCTTGGACGTGATTTGGTTCTCTTGGCGGTG 1549
 QY 1505 tgcctgacgtggcctcatcatcactttaaatgtgtgcttatggctaccgcggaatgcttgg 1564
 Db 1550 TGTGACAGTGGCCCTTCATCACCTTTTAAATGTTGTGCTTATGGCTACCGGAATTCCTTGG 1609
 QY 1565 ggaataaagggcgaggttaagaagcccaaaatccaagaccatgagaaagtgggtggga 1624
 Db 1610 GGAATAAAGCGCAGTTTAAAGAACCCCAAAATCCAAAGACCCATTTGAGAGTGGTGGGA 1669
 QY 1625 aataaggttaaaatttgaaccattcccttagtcatcattccaaacttgaaaaacagaaatcagtg 1684
 Db 1670 AATAAGGTAAAAATTTGAACCATTCCTTAGTCAATTTCCAAACTTGAAAAACAGAAATCAGTG 1729
 QY 1685 ttaaattcctttattcttattaaaggaataactttgcataaaatcaatcagccccagatg 1744
 Db 1730 TTAATTTCTATTTATTTAATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1789
 QY 1745 ctttaaaatctcttaataaaataatagactcgctagtcagtaagtaagatttgaat 1804
 Db 1790 CTTTAAAAATTTCTTAAATTAATAANTAGACTCGTAGTACGTAAGATATTTTGAAT 1849
 QY 1805 atgtatcgtgccccctccgtgcttcttgatcaggtgacatgctgcatttttcagaggac 1864
 Db 1850 ATGTATCTGTGCCCTCCCGTCTCTTGTATCAGGATGACATGTGCATTTTTCAGAGGAC 1909
 QY 1865 gtgcagaccggctggcatctagattactttctactctgaacatgctgctgttggga 1924
 Db 1910 GTGCAGACAGCGTGGCATCTAGATTACTTTTCTACTCTGAAACATGGCTGTTTGGGA 1969
 QY 1925 gtgcgggattcaaaaggtggtccccccgctgccccactgcaaatggcagttttaaactta 1984
 Db 1970 GTGCGGGATTCAAAGGTGTGCCACCGCTGCCCTACTGCAAAATGGCAGTTTAAATCTTA 2029
 QY 1985 tcttttggctctgcagatgggttgcattgac-cttaaccaataatggtcagtcctcat 2043
 Db 2030 TCTTTTGGCTTCTGCAGATGGTTGCAATGATCTCTTAACCAATAATGGTCACTCTCAT 2089
 QY 2044 ctctgctcttcataggtgccacctgtgtgttttaaaagaggaagcctttgtaccttt 2103
 Db 2090 CTCTCTCTTCTCTAGGTGCCACTTGTGTGTTTAAAGAGGGAAGCTTTGTACTCTTT 2149
 QY 2104 agagtgtaggtgaaatgaatgagttggtgagtgactgagacagcatatgatttctt 2163
 Db 2150 AGAGTGTAGTGAATGAATGAATGGCTTGGAGTGCACGTGAGACAGCATATGATTTTTT 2209
 QY 2164 gctttgggaaaaagaaatgactatgaaattgggtgtgtatttgagaagataatc 2223
 Db 2210 GCTTTGGGAAAAAGAAATGATCTATGAAATTTGGTGGGTGTATTTTGAGAAAGATAATC 2269
 QY 2224 attgcttatgtcaaatggagctgaatttgataaaaaaccccaaaacacagctatgagtgct 2283
 Db 2270 ATTGCTTATGTCAATGGAGCTGAAATTTGATAAAACCCCAAAATACAGCTATGAAGTGCT 2329
 QY 2284 gggcaagtttacttttttctgatttcttcaact 2320
 Db 2330 GGGCAAGTTACTTTTTTTTGTGATGTTTCTCTCAACT 2366

RESULT 6
 BC020971
 LOCUS

2423 bp mRNA linear PRI 22-JAN-2002

QY 1205 acaatgcaagcgcatgagactaagggagctggagtgaccctggaatgtctctggaatga 1264
 |||||
 Db 1220 ACAAAGCGCATGAGAGCTAAGGAGCTGGAGTGACCCCTGAATGTTCTGGAATGA 1279
 |||||
 QY 1265 ctctgaagatttaaaaaagctcttaaaagcagtcacatgaacaaagttaacagaga 1324
 |||||
 Db 1280 CTCTGAAGATTGAAATGCTCTAAAGCAGTCATCAATGACAAAGATTACAGAGAGA 1339
 |||||
 QY 1325 acatcatgagcctctccagccttcaaaagaccgccgggtggagcgctggacacctggccg 1384
 |||||
 Db 1340 ACATCATGCGCTCTCCAGCTTCAACAGCACCGCCGGTGGAGCGCTGGACCTGGCCG 1399
 |||||
 QY 1385 tgrtctggtgagtgattgataagcacaaagggcgccgcacacacctggagcccgacccc 1444
 |||||
 Db 1400 TGTTCTGGGTGGAGTTTGTGATGAGGCACAGGGCGGCCACACCTGCGCCGCCGACGCC 1459
 |||||
 QY 1445 acgacctcacctgggtaccagttaccattccttggacgtgattgttctctcttggccgtcg 1504
 |||||
 Db 1460 ACGACCTACCTGGTACCAGTACCATCTTGGACGTGATGGTCTTCTCTTGGCCGCTG 1519
 |||||
 QY 1505 tgcgtacagtgcccttcacaccttaaatgttgtcttattgctaccggaaatgcttg 1564
 |||||
 Db 1520 TGTGACAGTGCCCTTCATACCTTTAAATGTTGCTATGCTACCGGAAATGCTTGG 1579
 |||||
 QY 1565 ggaataaagggcgagtttaagaaagcccaaaatccaaagaccattgagaagtggtggga 1624
 |||||
 Db 1580 GGAATAAAGGGGAGTTAAGAAAGCCACAAATCCAAAGCCCATTTGAGAAGTGGGTGGGA 1639
 |||||
 QY 1625 aataaggtataatttgacacattcctcagtcattcctcaacttcaacttgaaacagaaatcagt 1684
 |||||
 Db 1640 AATAGGTAAATTTTGAACCATTCCTAGTCATTTTCCAACTGTAACACAGATCAAGT 1699
 |||||
 QY 1685 ttaaatcattttattcttataagaaatactttgcataaataatcagcccgagagt 1744
 |||||
 Db 1700 TTAATATTCATTTATTCTTATTAAGGAATACTTTGCATAAATAATCAGCCCCAGAGTG 1759
 |||||
 QY 1745 ctttaaaatctcttaataaataataagactcagtcagtaagataatttgaat 1804
 |||||
 Db 1760 CTTTAAATAATTTCTTTAAATTAATAATATAGCTCGCTAGTCAGTAAGATATTTGAAT 1819
 |||||
 QY 1805 atgtatgctgccccctccggtgtctttgatcaggatgacatgtgccattttttcagaggaac 1864
 |||||
 Db 1820 ATGTATGTGTCGCCCTCCGGTGTCTTTGATCATAGATGACATGTGCCATTTTTCAGAGGAC 1879
 |||||
 QY 1865 gtgcagacagctggcattctagattacttcttactctgaaacatgacctgtttggga 1924
 |||||
 Db 1880 GTGCAGACAGCTGGCATTCTAGATTACTTTTCTTACTCTGAACATGSCCTGTTGGGA 1939
 |||||
 QY 1925 gtgcggattcaaaaggttggtccaccgctgcccctactgcaaatgagcattgttaatttta 1984
 |||||
 Db 1940 GTGCGGATTCAAAAGGTGTGCCACCGCTGCCCTTACTGCAATGTCAGCTTTTAATCTTA 1999
 |||||
 QY 1985 tctttggtctctgagatggttgcaatgatccttaacaaataatggtcagtcctcctc 2044
 |||||
 Db 2000 TCTTTGGCTCTGAGATGTTGCAATGATCCCTTAACCAATAATGTCAGCTCCTCATC 2059
 |||||
 QY 2045 tctctctctctaataggtgccaacctgtgtgttttaagaaggaagcgtttgtacacctta 2104
 |||||
 Db 2060 TCTGCTGCTTCATAGGTGCCACCTGTGTGTTTTAAAGAGGGAAGCTTTGTACCTTTA 2119
 |||||
 QY 2105 gagtgtagtgaaatgaatgaatgctgtgagtgcaatgacatgagacagcatatgattcttg 2164
 |||||
 Db 2120 GAGTGTAGTGAATGAATGAATGGCTTGGAGTGCACACACACATATGATTTTTTG 2179
 |||||
 QY 2165 ctttgggaaaaaatgatgctatgaaatttggtgggtgtgtatttgagaagataatca 2224
 |||||
 Db 2180 CTTTGGGAAAAAAGATGATGCTATGAAATTTGGTGGTGTGATTGTGAGAAGATAATCA 2239
 |||||
 QY 2225 ttgcttatgtcaaatggagctggaatttgataaaaccccaaatcacagctatgaagtgtg 2284
 |||||
 Db 2240 TTGCTTATGTCAATGGAGCTGAATTTGATTAAAAACCCCAAAATACAGCTATGAAGTGTG 2299
 |||||

QY 2285 ggcaagtttaactttttctgtgatgttttctactaacact 2320
 |||||
 Db 2300 GGCAAGTTTACTTTTTTTTGTGATGTTTCTCTCAACT 2335
 |||||
 RESULT 5
 AF462268 2394 bp mRNA linear PRI 07-FEB-2002
 LOCUS Homo sapiens UDP-glucuronosyltransferase UGT1A8*1 (UGT1A8) mRNA,
 UGT1A8*1 allele, complete cds.
 AF462268
 VERSION AF462268.1 GI:18568218
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2394)
 Huang, Y.-H., Nguyen, N., Geske, D. and Tukey, R.
 Identification and functional characterization of UGT1A8*1,
 UGT1A8*2 and UGT1A8*3
 Unpublished
 JOURNAL 2 (bases 1 to 2394)
 Huang, Y.-H., Nguyen, N., Geske, D. and Tukey, R.
 Direct Submission
 AUTHORS Submitted (21-DEC-2001) Pharmacology, UCSD, 9500 Gilman Drive, La
 Jolla, CA 92093, USA
 JOURNAL
 TITLE Location/Qualifiers
 FEATURES
 source 1..2394
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="colon"
 gene 1..2394
 /gene="UGT1A8"
 /allele="UGT1A8*1"
 64..1656
 /gene="UGT1A8"
 /codon_start=1
 /product="UDP-glucuronosyltransferase UGT1A8*1"
 /protein_id="A175964.1"
 /db_xref="GI:18568219"
 /translation="MARTGWTSPIPLCVSILLTCTGFABAGKLLVPMDSGSHWTFMQSV
 VERLILRGHEVVVPMPEVSQWGLSKNCTVKTYSYTLDELDFREMFDAQWRAQV
 RSLFLSSSNGFFNLFHSRSLFNDRKLVEYLKESDFDAVDFDFDACCGLIVAKY
 FSLPSVFAAGIACHYLEEAGCPAPLSPRILIGFSADMTKFRVRNIMHLEHL
 FCQVSKNALEIASEILQTPVATDLYSHTSIWLLRTDFLDYPKPMPNMIFIGIN
 CQCKPLPMEFEAYINASGEHGIWFSIGSWSEIPEKKAMADALGIPQTVLMRY
 TCTPNSLANNTILVWLPPNDLLGHMTAFITHAGSHGVYESICNGVPMVMPPLFG
 DQMDNKRMTKGACVTLNVLFTSEDLNALKAVINDKSKYKINMLSLHDKRPE
 PLDIAVFWFEFVRHKHGAHLRPAADHDLTWYQYHSLDVIGFLAVLVITVAFTFKCCA
 YGYRKCGLGKGRVKRAHKSHTH"
 BASE COUNT 624 a 513 c 555 g 702 t
 ORIGIN
 Query Match 95.1%; Score 2207.4; DB 9; Length 2394;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 2255; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
 QY 5 gctgcagttctctgatggcttgacaggggtggaccagccccctctctctatgtgtgtc 64
 |||||
 Db 50 GCTGCAGTTCTCTCATGCTGCGCACAGGCTGCACGCCCATTCCTCTATGTTTCTC 109
 |||||
 QY 65 tgcgtgacctctggtgttcgagagagaggaagactactggaagtcagccatggaatgga 124
 |||||
 Db 110 TGCTGCTGACCTGTGTGGCTTCTCTGAGCGAGGAAGCTCTGCTAGTAGTGCCTATGATGGGA 169
 |||||
 QY 125 gccactggttcaccatgaggtgcggtggtggagaaactcattctcagggggcaggtg 184
 |||||
 Db 170 GTCAGTGTTCACCATGCAGTGTGCTGGAGAAACTTATCCTCAGGGGCGCATGAGTGG 229
 |||||
 QY 185 ttatgactatgcacagatgagttgcaactggagagctcactgaattgcacagtgaga 244
 |||||

QY	2161	cttgccttggggaagaaatgactgataaattggtgggtggtgatttattgagaagata	2220	Db	80	TGCTGCTGACCTGTGGCTTTGCTGAGGACAGGGAAGCTGCTGTGTAGTGCCTCATGATGGGA	139	
Db	2161	CTTGCTTTGGGGAAGAAATGATGCTATGAATTTGGTGGTGTCTATTTGAGAAGATA	2220	QY	125	gcaactggttcacacatgaagtcggttggtggagaactcattctcaggggggcatgaagtggtg	184	
QY	2221	atcattgcttatgtcaaatggagctgaatttgataaaaccccaaaatacacagctataa	2280	Db	140	GTCATGCTTCCATCTCCATGCTCGGTGGTGGAGAACTTATCTCTCAGGGGGCATGAGGTGG	199	
Db	2221	ATCATGCTTATGTCGAATGGAGCTGAATTTGATMAAACCCCAAAATACAGCTATGAAGT	2280	QY	185	tgtgagtcactgcccagaggtgagttggcaactgggaagatcactgaattgacagtgaa	244	
QY	2281	gctgggcaagttaactttttctgattgtttccataaact	2320	Db	200	TTGTAGTCAITGCCAGAGGTGAGTTGGCAACTGGGAAATCACTGAATTTGCACAGTGAAGA	259	
Db	2281	GCTGGCAAGTTACTTTTTTCTGATGTTTCTTCTACAACT	2320	QY	245	cttattcaactcatatccctggagatcctggaacgggagttcaaggcttttgcacatg	304	
RESULT	4			Db	260	CTTACTCAACTCATACATCTCGAGAGATCTGGACCGGAATTCATGATTTGCCGATG	319	
LOCUS	AF462267	2363 bp	mRNA	linear	QY	305	ctcaatgggaagcacgaagtacgaagataatatctctattatgggtttcatacaaatgaca	364
DEFINITION	Homo sapiens UDP-glucuronosyltransferase UGT1A8*2 (UGT1A8) mRNA.			Db	320	CTCAATGGGAAGACACAAGTACGAAGCTTTGTTTCTCTATTTCTGAGTTTCATCCAAATGGTT	379	
ACCESSION	AF462267			QY	365	ttttgacttatttttcaaatgcaggagttgttttaagacacaaaaattagtagaat	424	
VERSION	AF462267.1	GI:18568216		Db	380	TTTTTAACTTATTTTTCGCATTTCAGGAGTTGTTTAAATGACCGAAATTAAGTAAAT	439	
KEYWORDS	human.			QY	425	acttaagagagagttcttttgatgacagtggttcttcgatccttttgataaacttgggttaa	484	
ORGANISM	Homo sapiens			Db	440	ACTTAAAGAGAGATGCTTTTGTATGCGGTGTTTCTTGTATCTCTTTTGTATGCTTGTGCTTAA	499	
REFERENCE	1 (bases 1 to 2363)			QY	485	ttgttgccaaatatttctccctccctccgtggtcttcgcaggagggaatactttgcca	544	
AUTHORS	Huang, Y.-H., Geske, D., Nguyen, N. and Tukey, R.			Db	500	TTGTTGGCCAAATATTTCTCCCTCCCTCTCTGCTCTCTCCGACAGGGGAATAGTTGCCACT	559	
TITLE	Identification and functional characterization of UGT1A8*1, UGT1A8*2 and UGT1A8*3			QY	545	atcttgaagaggtgcacagtgccctgctcctctctctctctctctctctctctctctct	604	
JOURNAL	2 (bases 1 to 2363)			Db	560	ATCTTGAAGAGGTGGACAGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	619	
REFERENCE	Huang, Y.-H., Geske, D., Nguyen, N. and Tukey, R.			QY	605	ggttctcagatgctcagcttctcaagagagagtagcaggaacacacatcatgcacttggagg	664	
AUTHORS	Direct Submission			Db	620	GGTCTCAGATGCCATGACTTTTCAAGGAGAGAGTACGGAACACATCATGCATCTGGAGG	679	
TITLE	Submitted (21-DEC-2001) Pharmacology, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA			QY	665	aacatttatatagcaccgcttttttcaaaaatgccctagaaaatagcctctgaaattctcc	724	
JOURNAL	Location/Qualifiers			Db	680	AACATTTATTTTGGCAGTATTTTTCAAAATGCCCTAGAAATAGCTCTGAAATTTCTCC	739	
FEATURES	1. .2363			QY	725	aaacacctgttcaggagtagatctctacagccacacatacaatttggttgttgcgaacagg	784	
Source	/organism="Homo sapiens"			Db	740	AAACACCTGCACAGCATATGATCTCTACAGCCACACATCAATTTTGGTGTGTCGAACAG	799	
gene	/db_xref="taxon:9606"			QY	785	actttgtttggactatcccaaacccgtgtagcccaacatgatctctctctctctctctct	844	
CDS	/tissue_type="colon"			Db	800	ACTTTGTTTGGACTATCCCAACCCGTGATGCCCAATATGATCTTCATTTGGTGGTATCA	859	
	1. .2363			QY	845	actgccaatcagggaagcgttgctctatggaattggaagcctacataaattgctctcgag	904	
	/gene="UGT1A8"			Db	860	ACTGCCATCGGGAAGACCCATTTGCCCTATGGAATTTTGAAGCCTACATTAATGCTCTCGAG	919	
	/allele="UGT1A8*2"			QY	905	aacatgaaattggttttctctctctctctctctctctctctctctctctctctctctct	964	
	34..1626			Db	920	AACATGGAATTTGTGTTTCTCTCTTTGGGATCAATGGTCTCAGAAATTTCCAGAGAAGAAG	979	
	/gene="UGT1A8"			QY	965	ctatggcaattgctgatgctttgggcaaaatccctcagacagctccttgccggtacactg	1024	
	/codon_start=1			Db	980	CTATGGCAATTTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCTCTGTGCGGTGACACTG	1039	
	/product="UDP-glucuronosyltransferase UGT1A8*2"			QY	1025	gaacccgaccatcgaatcttgcgaacacacgatacttgtttaagtggtctaccaccaaacg	1084	
	/protein_id="AAL75963.1"			Db	1040	GAACCCGACATCTCGAATCTTGGCAACCAACAGATACTTGTAAAGTGGCTACCCCAAAACG	1099	
	/db_xref="GI:18568217"			QY	1085	atctgcttggttcacccgagacccgtgctctttatcaccatgcttggttcccatggtgttt	1144	
	/translation="MARTGWTSPILCVSLILITCGFAEAGKLLVVPMDGSHWETQSV			Db	1100	ATCTGCTTGTCTACCCGATGACCCGCTGCTTTATCACCATGCTGTTCCCATGTTGTTT	1159	
	VEKLIRGHEVVMPEVSMQGLKSLNCTVKTYSTYLTEDLDFEPMDFADQWKAQV			QY	1145	atgaaacacatgcaatggcttcccatgtagatgctcttctctctctctctctctctctct	1204	
	RSLSFLSSNGCFNLPSHCRSLFNDRKLYELAKSSFDVFLDFDPDAGLIVAKY			Db	1160	ATGAAAGCAATGCAATGGCGTTCCATGTTGATGCTGCTGTTGTTGTTGTTGTTGTTGTTG	1219	
	FEPSYFARGICGCHLEGAQC PAPUSYVRILLGFSAMTFKRVNRNIMHLEHL							
	FCQFSKALIEASEILQTPVAVDLSYTSIWLRTDEVLDYKPKVPMNPIF IGIN							
	CHQRPUPMEFEAYINASGEHIVVFSIGSMVSEIPEKKAMAIDALCKIPQTVLWRY							
	TTSPSLNANNITLVKWLQNDLLHPWTRAFITHAGSHGVYESICNGVPMVWMPFLF							
	QDMDNAKMETKAGVTLNVLMTSEDLNALKAVINDKSYKENIMRLSLIHDRPVE							
	PDLAVFWVEFVNRHKGAPHLRPAHDLITWYQVHSLDVLVGLLAVLTVAFITFKCA							
	YGYRCLGKKGKRVKKAHKSHT"							
BASE COUNT	617 a 503 c 551 g 692 t							
ORIGIN								
Query Match	95.6%	Score	2218.4	DB	9	Length	2363	
Best Local Similarity	97.4%	Pred. No.	0					
Matches	2255	Conservative	0	Mismatches	61	Indels	0	
Gaps	0							
QY	5	gctgcagttctctgagcttgacagagggtgacacgcccctctctctctctctctctctct	64					
Db	20	GCTGCAGTTCTCTCAGTGGCTCGCACAGGGTGGACACGCCATTCCTCCCTATGTTCTC	79					
QY	65	tgctgctgacctgtggtcttccgagcagcaggaagcactggttagtgcacatggatggga	124					

Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	ctcagctgcagttctctgattggttcacaggttgacagccccccttctctatgtgtg	60
Db	1	CTCAGCTGCAGTTCCTGATGGCTGCACAGGGTGGACAGCCCCCTCTATGTGTG	60
Qy	61	tgtctctgtgaacctgtgctttgccagagcaggggaagactactagtagtcccatggt	120
Db	61	TGCTCTGCTGACCTGTGGCTTTGCCAGGCGAGGAAGCTACTGGTAGTCCCATGGAT	120
Qy	121	ggagacactggttcacacatgaggtcggttggtagaactcattctcaggggcatgag	180
Db	121	GGGAGCCACTGGTTCACCATGAGTGGTGGTGGAGAACTCATCTCAGGGGCGATGAG	180
Qy	181	gtggttgtagtccagaggttgagttggaactgggaagatcaactgaattgcaagtg	240
Db	181	GTGGTGTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATGGCACAGT	240
Qy	241	aagacttattcaacttatataccctggaggatctggaccgggagttcaaggctttggc	300
Db	241	AAGACTTATTCAACTTCATATACCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC	300
Qy	301	catgctcaatgaaagacaaagtaagaagtatatctctatttaattgggttcatacaat	360
Db	301	CATGCTCAATGGAAGCACAAGTACGAAGTATATTCTCTATTAATGGGTTCATACAA	360
Qy	361	gaacttttgacttattttttcaaatgtcagaggttgttttaagacaaaataatagta	420
Db	361	GACATTTTGTGACTTATTTTTCAAATTCAGGAGTTGTTTAAAGACAAAAAATTAGTA	420
Qy	421	gaatacttaagagaggttcttttgatgcagtggtttctcgatcccttttgataactggc	480
Db	421	GAATACTTAAAGAGAGTCTTTTGTGATGCAGTGTCTTCGATCCTTTTGTATTAATGTGGC	480
Qy	481	ttaattgtgcacaaatattctccctccctccgttgcgttctgccaggggaaactcttgc	540
Db	481	TTAATTTGTTGCCAAATATTTCTCCCTCCCTCCGTGTCTTGCCAGGGGAATACTTTGC	540
Qy	541	caatatctgaagaagtgacagtgccctactctctctctctctctctctctctctctc	600
Db	541	CACATCTCTGAAGAAGTGCACAGTGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	600
Qy	601	ttaggtttctcagatgcacatgactttcaaggagagagtacggaacacacacacacac	660
Db	601	TTAGGGTCTCAGATGCCTGACTTTTCAAGGAGAGAGTACGGAACACATCATGCACTTG	660
Qy	661	gaggaacattatgaaccctgttttcaaaaatgccctagaatagcctctgaatt	720
Db	661	GAGGAACATTTTATATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAAT	720
Qy	721	ctcaaacacctgttaaggagtagatctctacagccacacacacacacacacacacac	780
Db	721	CTCCAAACACACCTGTTACGGAGTATGATCTTACAGCCACACATCAATTTTGGTGTGCGA	780
Qy	781	acagacttttttgactatcccaaacccgtgatgcccaacatgactctctctctctctct	840
Db	781	ACGACTTTGTTTGGACTATCCCAACCCGTTGATGCCCAACATGATCTTCTATGTGGGT	840
Qy	841	ataactgpcatcagggaaagccgttgccctatggaaattgaaacctacataatgctct	900
Db	841	ATCAACTGCCATCAGGAAAGCCGTTTGCCCTATGGAATTTGAAAGCCCTACATTAATGCTCT	900
Qy	901	ggagaacatggaattgtgttttctcttgggaatcaatgctcagaataatccagagaag	960
Db	901	GGAGAACATGGAAATGTGGTTTTCTTTTGGGATTCATGGTCTCAGAAAATTCAGAGAG	960
Qy	961	aaagctatggcaattgtgtatgctttgggcaaaaacctcagacagctcgttggcggtac	1020
Db	961	AAAGCTATGGCAATTTGCTGATGCTTTTGGGCAAAATCCCTCAGACAGTCTCCTGTGGCGGTAC	1020
Qy	1021	actggaacccgacctcaatcttgcgaacacacagatacttgttaagtggtcaccacaa	1080
Db	1021	ACTGGAACCCGACCATCGAATCTTTGGCAACACACAGATACTTGTTAAGTGGCTACCCAA	1080
Qy	1081	aacgatctgtgtgaccccgatgacccgtgctcttataccaccatgctggttcccatggt	1140
Db	1081	AACGATCTGCTTGGTACCCGATGACCCGTCCTTATACCCATGCTGTGTTCCATGGT	1140
Qy	1141	gtttatgaaacatgatgaatggttcccatgctgatgaccccttgttgggtgacag	1200
Db	1141	GTTTATGAAACATATGCCAATGGCGTTCATGGTGTGATGATGATGATGATGATGATG	1200
Qy	1201	atggacaatgcgaagcgatgagagtaaggagctggtgagctgaaatgttctggaa	1260
Db	1201	ATGGACAATGCAGGAGCGATGAGACTAAGGGAGCTGGAGTGACCTGAAATGTTCTGGAA	1260
Qy	1261	atgactctgaagatttadgaataatgcttaaaacagatcatcaatgacaaagttaacaag	1320
Db	1261	ATGACTCTGAAGATTTAGAAAATGCTCTAAAACAGTCAATCAATGACAAAAGTTACAG	1320
Qy	1321	gagaacatcatgcctctccagccttcaagaaggacccgctggagcctggagacctg	1380
Db	1321	GAGAACATCATGCGCTCTCCAGCCTTACAGGACCGCCGCTGGAGCCCTGGACCTG	1380
Qy	1381	gccgtgttctggtgaggttctgtatgagggcacaaaggcgccacacactgcgcccgca	1440
Db	1381	GCGTGTCTGGTGGAGTTTGTGATGAGGCACAAAGGCGCGCCACACTGCGCCCGCA	1440
Qy	1441	gccacgacctcaactggtaccagttaccattccttgggaogtgaatggttctctctggcc	1500
Db	1441	GCCACGACCTCAGCTGTTACCAAGTACCAATTCCTTGGACGTGATTTGGTTCTCTTGGCC	1500
Qy	1501	gtcgtctgacagtggtccttcaacccctttaaigtgtgtcttatggtaccggaaatgc	1560
Db	1501	GTGCTGTGACAGTGGCTTTCATCACCCTTAAATGTTGTCTTATGGCTACCGGAAATGC	1560
Qy	1561	ttggggaaaaagggcgagtaagaagaacccacaaatccaagaacccattgagaagtggt	1620
Db	1561	TTGGGAAAAAGGGCGAGTTAAGAAGCCCAAAATCCAAGACCCATGTAGAGAGTGGGT	1620
Qy	1621	ggaaataaggttaaaatttgaacacattccctagtcatttccaaacttgaaaaacagatc	1680
Db	1621	GGAAATAAGGTAAAAATTTGAACCATTTCCCTAGTCATATTTCCAAACTTGAACACAGATC	1680
Qy	1681	agtgtaataatcatttattcttatttaaggaaataactttgacataataataacacccag	1740
Db	1681	AGTGTAAATCATTTTATTCTTATTAAAGAAATACCTTGCATAAATTAATACGCCCCAG	1740
Qy	1741	agtgcttttaaaaaattctcttaataaaaaataatagactcgttagtcaagtaaaatatt	1800
Db	1741	AGTGTCTTAAAAAATTTCTTTAAATAAAAAATAATAGACTCGCTAGTCAATAAGATATT	1800
Qy	1801	gaatatgtatcgtgcccccctccggtgtcttttgatcagatgacatgtgccattttcaga	1860
Db	1801	GAATATGTATCGTGCCCCCTCCGGTGTCTTGTATCAGAGTGCATGTGCCATTTTTCAGA	1860
Qy	1861	ggacgtgcagacagcgtggcattcttagattactttcttactctgaacatgacctgttt	1920
Db	1861	GGACGTGCAGACAGCTGGCATTTCTAGATTACTTTCTTACTCTGAAACATGGCTGTTT	1920
Qy	1921	ggagtgaggattcaagtggtccacccgctcccactactgcaaatggcagttttaat	1980
Db	1921	GGGAGTGGGGATTCAAGAGTGTCCACCGCTCCCTACTGCAAAATGGCAGTTTTAAT	1980
Qy	1981	cttatcttttggcttctgcagatggttgcaattgactcttaacccaataatggtcagtcct	2040
Db	1981	CTTATCTTTTGGCTTCTGCAGATGGTTTGCAATTGATCCTTAACCAATATGTCAGTCT	2040
Qy	2041	catctctgctctgcttcaataggtgcacactgtgtgtttaaagaaggaagctttgtacc	2100
Db	2041	CATCTCTGCTGCTCATAGGTGCGCCACTTGTGTGTGTAAAGAAAGGAGCTTTGTACC	2100
Qy	2101	tttagagtgtagtgaaatgaaatgaatggtctgagtgactgagaaacagcatatgattt	2160
Db	2101	TTTAGAGTGTAGTGAATGAATGAATGGCTTGGAGTGCATGTGAGAACAGCATATGATT	2160

Db	1021	ACTGGAACCCGACCACATCGAATCTTCGGAAACAACACGATACTGTTAAGTGGCTACCCCAA	108
QY	1081	aacgactctgttggctaccccgatgaccctgccttcttaaccactgctggttcccacgtt	1140
Db	1081	AACGATCTGCTGTGCTACCCGATGACCGCTGCCTTTATCACCCATGCTGCTTCCCATGGT	1140
QY	1141	gtttatgaagcatatgcgaatggcttcccatggtgtagtgatgccttgtttggtgatcaag	1200
Db	1141	GTTTATGAAAGCATATGCAATGGCTTCCTCATGGTGATGATGCCCTGTGTTGGTGATCAG	1200
QY	1201	atg9gacaaatgcgaagcgcatggagacataaggagctggagtgaccctggaatgttctggaa	1260
Db	1201	ATGGACAATGTCGAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCCTGAATGTTCTGGAA	1260
QY	1261	atgacttctgaagatttagaaatgctctctaaagcagtcacatgacaaaaagttacaag	1320
Db	1261	ATGACTTCTGAAGATTGTAGAAATGCTCTAAAGCAGTCATCAATGACAAAAAGTTTACAAG	1320
QY	1321	gagaaacatgcgctctctccagcttccaaagaccctggctggagcgcgtggaactgtgacctg	1380
Db	1321	GAGAAACATCATGCGCTCTCCAGCTTCCAAAGGACCCCGGTGGAGCCGCTGGACCTG	1380
QY	1381	gcctgtcttgggtggagttttagatgagcacaagggcgcgcaacacctgcgcccgcga	1440
Db	1381	GCCGTGTTCTGGGTGGAGTTGTGTATGAGGCACAGGGCGGCCACACTGCGGCCCGCA	1440
QY	1441	gccacgacctccactggttaccagtaaccttctggagctgattggttctcttcttggcc	1500
Db	1441	GCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTCTCTTGGCC	1500
QY	1501	gtcgtctgcagtgggccttcacactttaaattgtcttatgctacgcgaagaatgc	1560
Db	1501	GTGCTGCTGCAGGTGGCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGC	1560
QY	1561	tgggggaaaaaggggcgagtgaagaagcccaataccaagaccctatgagaagtgggt	1620
Db	1561	TTGGGAAAANAAGGGCGAGTTAAGAAAGCCCAATAATCAAGACCCATTGAGAAGTGGGT	1620
QY	1621	gggaataaaggtaaaatttggacaattccctagtcatttccaaacttgaaacacagaatc	1680
Db	1621	GGCAAAATAGGTAAATTTTGAACATTCCTTAGTCATTTCCAAACTTGAAACAGAAATC	1680
QY	1681	agtgtaaatcattttattcttatttaagaaatacttgcataataatcagccccag	1740
Db	1681	AGTGTTAAATTTCATTTTATTTCTTATTAAAGAAATACTTTGCATAAATTAATCAGCCCCAG	1740
QY	1741	agtggtttbaaaaaattctttaaataaaaaataatgagctcgctagtcagtaagaatat	1800
Db	1741	AGTGTCTTTAAAAAATTCCTTTAAATAAAAAATAATAGACTCGCTAGTCAGTAAGATATTT	1800
QY	1801	gaatgtatcgtgccccctccggtgctttagatcagatgacatgctccatttttcaga	1860
Db	1801	GAATATGTATCGTGCCCCCTCCGGTGCTTTGATCAGATGATGATGCCATTTTTCAGA	1860
QY	1861	ggaagtgcagacggctggcatttagattacttttcttactctgaacatggcctgttt	1920
Db	1861	GGAGCTGCAGACAGCTGGCATCTCAGATTACTTTTCTTACTGTGAACATGCGCTGTATT	1920
QY	1921	gggaagtgcgggaattcaaaagtgtgtccacccgctgcacctactgcgaatggcagtttcaat	1980
Db	1921	GGGAU TGCGGGAATTCAAAGGTGTCCACCGTCCCTACTGCANATGGCAGTTTTAT	1980
QY	1981	cttaacttttggcttctgcagatggttgcgaatgtgactctaaaccaataatggctcact	2040
Db	1981	CTTACTTTTGGCTCTGCAGATGGTTGCAATGTATCTTAAACCAATAATGGTCAGTCT	2040
QY	2041	catcctgtcctccttcataagtgccacctgtgtttttaaagaagggaactgttgacc	2100
Db	2041	CATCTCTGCTCTTCATAGTGTGCCACCTTGTGTGTTTAAAGAGGGAACTTTGTACC	2100
QY	2101	ttaatgagttagtgaaatgaatgaatgcgtgcgtgcactgcactgcagaacagatatgatt	2160
Db	2101	TTTAGAGTGTAGGTGAAATGAATCAATGCTTGGAGTGCACTGAGAACAGCATATGATTT	2160

QY	2161	cttgctcttgggaaaaagaatgatgctatgaataatggtgggtggtattgattgagaagata	2220
Db	2161	CTTGGCTTTTGGGGAAAAAGAAATGATGCTATGAAATTGGTGGGTGGTGTATTTTGAAGATA	2220
QY	2221	atcattgctctatgtcctaaatggagctgaatttgataaaaccccaaatcacgctatgaagt	2280
Db	2221	ATCATTGCTTATGTGCAATGGAGCTGAATTTGATAAAACCCCAATACACCTATGAAGT	2280
QY	2281	gctgggcaagtctactttttcttctgattgttctctacaact	2320
Db	2281	GCTGGGCAAGTCTACTTTTCTTCTGATGTTCTCTCAACT	2320
RESULT 3			
LOCUS	AF056188	2320 bp	mRNA linear PRI 13-AUG-1998
DEFINITION	Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA, complete cds.		
ACCESSION	AF056188		
VERSION	AF056188.1 GI:3414797		
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1. (bases 1 to 2320)		
	Ciotti,M., Potter,C. and Owens,I.S.		
	Human phenol metabolizing UDP-glucuronosyltransferase		
REFERENCE	Unpublished		
	2. (bases 1 to 2320)		
	Ciotti,M., Potter,C. and Owens,I.S.		
	Direct Submission		
REFERENCE	Submitted (30-MAR-1998) Heritable Disorders Branch, NICHD, 9000 Rockville Pike, Bethesda, MD 20892, USA		
	3. (bases 1 to 2320)		
	Ciotti,M., Potter,C. and Owens,I.S.		
	Direct Submission		
REFERENCE	Submitted (12-AUG-1998) Heritable Disorders Branch, NICHD, 9000 Rockville Pike, Bethesda, MD 20892, USA		
	TITLE		
	Sequence update by submitter		
	On Aug 13, 1998 this sequence version replaced gi:3025895.		
REMARK	Location/Qualifiers		
COMMENT	1..2320		
FEATURES	/organism="Homo sapiens"		
source	/db_xref="taxon:9606"		
	/chromosome="2"		
	/map="2q37"		
	/tissue_type="liver"		
gene	1..2320		
	/gene="UGT1A9"		
CDS	19..1611		
	/gene="UGT1A9"		
	/codon_start=1		
	/product="UDP-glucuronosyltransferase 1A9"		
	/protein_id="AAC31425.1"		
	/db_xref="GI:3025896"		
K13LRLHVVVWVPEVSMQGRSLNCTVKYTSYTLDDREFKAFACQAKVAY RGIYSLMGNDIDFLFNSRSLCKDKLYEVLKESFADFVLPDNCGLIVAKY FSLPSVFAHGILCHYLFEEGAQCAPLSVYPRILLGSFDMATFKERVNHIHLHEHL LCHEFNKALEISELITOPVTEYDLYSHTSIWLIRTDVLDYKPYWPNKITEGGIN CHQGLDPHEEYAINASGEGHIVFSLGSWSEIPEKKAIAIADLGIKIQTVLWRY TGTRFSNLANITLWKLQNDLLGHPHTFATITTHAGSHGYVESICNGVPMWMPRLFE DQMDNKNMETKAGVITLVNLEMTSBDLENAKAVINDKSYKESINMLRSLSLHKDRPFE PLDLAVWFVEFVNRHKGAPHLNLSLSDVIGFLVAVLTVAEIFTCCKA			
BASE COUNT	608 a	495 c	547 g 670 t
ORIGIN	YGYRCKLGKRGYKKAHRSKTH"		

```

|||||
Db 1861 GGACGTGCAGACAGCTGGCATCTAGATTACTTTCTTACTCTGAAACATGGCTGCTTT 1920
Qy 1921 ggagtgccggattcaagtggtccaccctccctactacgaaatggcagtttaat 1980
Db 1921 GGAGTGCAGGATCAAGGTGTCACCGCTGCCCTACTGCAAAATGGCAGTTTAAT 1980
Qy 1981 ctatcttttgctctgcagatggttgcaattgatccttaaccataatgctcagtcct 2040
Db 1981 CTTATCTTTGGCTTCTGCAGATGTTGCAATTCATCTTAACCAATAATGGTCAGTCT 2040
Qy 2041 catctctgctctcatagtgccaccttgctgtgtttaagaagaggaaacttgtaac 2100
Db 2041 CATCTCTGCTCTCTCATAGTGCACCTTGTGTGTTAAGAGGGAAGCTTTGTAAC 2100
Qy 2101 tttagagtgagtgaaatgaatgaatggcttgagtgagtgagtgagtgagtgagtgag 2160
Db 2101 TTTAGAGTGTAGTGAATGAATGAATGGCTTGGAGTGCAGTGCAGAACAGCATATGAT 2160
Qy 2161 ctgcttgagggaagaaatgatctatgaattggtgggtggtgtatttgagaagata 2220
Db 2161 CTGCTTTGGGAAAGAAATGATGCTATGAATGGTGGTGGTGTATTTGAGAAGATA 2220
Qy 2221 atcattgcttattgaatggagtgaaatttgataaaaaaaccaatacagctatgaagt 2280
Db 2221 ATCATTTGCTTTATGTCGAATGGAGTGAATTTGATAAAAAACCAATAACAGCTATGA 2280
Qy 2281 gctggcgaagtttactttttctctgatgtttcctcacaact 2320
Db 2281 GCTGGCGCAAGTTACTTTTTTCTGATGTTCTTCACTCAACT 2320

RESULT 2
AX286764
LOCUS AX286764 2320 bp DNA linear PAT 21-NOV-2001
DEFINITION Sequence 3 from Patent WO0180896..
ACCESSION AX286764
VERSION AX286764.1 GI:17048799
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Ratkin,M.J., Innocenti,F. and Iyer,L.
Flavopiridol drug combinations and methods with reduced side
effects
JOURNAL Patent: WO 0180896-A 3 01-NOV-2001;
ARCH DEVELOPMENT CORPORATION (US)
FEATURES
source
1..2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
19..1611
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD12275.1"
/db_xref="GI:17048800"
/translation="MACITGWSPLPCLVCLLLTCGFAEGKLLVVPMDSGSHWFTWRSV
VEKILRGHEVVVPEVSWOLGRSLNCTKTSYSLTDLDRFKFAFHAWKQV
RSIYSLMGSYNDIFDLFFSNCRSLFKVLEKSESDAVFLDFPDCGLIVAKY
FSLPSVFARGILCHYLEEQAQCAPLSYVPRILIGSDAMTFKVRVNHIMHLEHL
LCRRFKNALEIASLEIQTPTVEYDLSHTSIWLLRTDFVLDYKPKVMPNMFIGGIN
CHQCKPLUPMEFAYINASGEHGVIVFSLGSMVSEIPEKKAMATADALGXIPOTVLWRY
TGTPRSLNANTLVKWPONDLLGHPTWRAFIHAGSHGVIESICNGVPMVMPFLFG
DOMDNARFKETKAGVTLNVLMTSEDLNALKAVINDKSKENTIMRLISLHKDRPVE
PILDNAFWVEFVNRKGAHPLRPAADHLTWYQYHSLDVLGELLAVLTVAFITFKCCA
FYGRKLGKGRVKKRAHKSHTH"
BASE COUNT 608 a 495 c 547 g 670 t
ORIGIN
|||||

```

Query Match 100.0%; Score 2320; DB 6; Length 2320;

```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ctccagctcagattctctgatgcttgccagggtagcagccccccttctctatgttg 60
Db 1 CTAGCTGCACTTCTCTGATGGCTTGACAGGGGAGGACCCCTCTCTATGTGTG 60
Qy 61 tgcctgctgagccttgctgcttccgagggcagggaaactactgtagtgcacatg 120
Db 61 TGTCTGCTGCTGACCTGTGCTTTCGAGGCGAGGAAAGCTACTGCTAGTGCCT 120
Qy 121 gggagccactggttcaccatcagagtgctgctgagaaactcatctcaggggcatg 180
Db 121 GGAGGCACTGTTTCACTATGAGCTGGTGGTGGAGAACTCACTGAATTCACAG 180
Qy 181 gtggtttagtcagtcagagtgagttggcaactgggaagatcactgaattgcacagt 240
Db 181 GTGGTGTAGTCAATGCCAGAGGTGAGTTGGCAACTGGGGAAGATCACTGAATTC 240
Qy 241 aagacttattcaactctcatataccctggagatctggaccgggagttcaaggtctt 300
Db 241 AAGACTTATTCACTTCAATATACCTGGAGGATCTGGACGGGAGTTCAGGCTTT 300
Qy 301 catgctcaatgggaagcacaagtagcagaagtatatattctcttaattgggttaca 360
Db 301 CATGCTCAATGGGAAGCACAAAGTACGAAGTATATATTTCTTATTAATGGGTCA 360
Qy 361 gacattttgacttattttttcaaatggcaggagttgttttaagacaaaaattag 420
Db 361 GACATTTTGTACTTATTTTTCATAATTCGAGAGTTTGTGTTTAAAGACAAAAA 420
Qy 421 gaatactaaaggagagctcttttgatcagtgcttctcgatcctttgataactgtg 480
Db 421 GAATATCTAAAGGAGAGTCTTTTATGATGAGTGTCTCGATCTCTTGTATTAAC 480
Qy 481 ttaattgttgccaaattttctccctccctccctggtctctccagggggaatact 540
Db 481 TTAATTTGTGCAAAATATTTCTCCCTCCCTCCCTGCTGCTTCCGCCAGGGAA 540
Qy 541 cactatctgaagaagtgacagctcctcctcctcctcctcctcctcctcctcctc 600
Db 541 CACTATCTTGAAGAAGGTGCACAGTGCCTGCTCTCTCTCTTCTATGTGCCAGAA 600
Qy 601 ttagggttctcagatgccatgacttccaagagagagtagcgaacacatcatcact 660
Db 601 TTAGGCTTCTCAGATGCCATGACTTTCAAGGAGAGAGTAGCGAAACACATCATC 660
Qy 661 gaggaacatttatatgccaccgtttttcaaaaaatgccctagaataagcctctga 720
Db 661 GAGGAACATTATTTATGCCACCGTTTTTTCAAAATGCCCTAGAAATAGCCCTG 720
Qy 721 ctcaaacactgttcaggagtagtctctacagccacacatcaattgtgttgctg 780
Db 721 CTCAAACACCTGTACGGAGTAGTATCTCTACAGCCACACATCAATTTGGTTG 780
Qy 781 acggacttgtttggactatcccaaacccgtgagcccaacatgatctctcatgtt 840
Db 781 ACGGACTTTGTTTGGACTATCCCAACCCGTGATGCCCAACATGATCTCTCAT 840
Qy 841 atcaactgccatcaggaagacgcgttgcttatggaatttgaagcctacattaaagt 900
Db 841 ATCAACTGCCATCAGGAAAGCCGTTGGCTTATGGAAATTTGAAGCCPACATAA 900
Qy 901 ggaagacatggaattgtggttttctctctttgggatacaatgggtctcagaata 960
Db 901 GGAGAACATGGAATTTGTTGTTTCTCTTTGGGATCAATGTTCTCAGAAATTC 960
Qy 961 aagctatgcaattgctgatgctttgggcaaaaatccctcagacagctcgtggcgt 1020
Db 961 AAAGCTATGGCAATTTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCTCT 1020
Qy 1021 actggaaacccgaccatcgaattcttcgaacaacacagatactgttaagtggtgc 1080
Db 1021 actggaaacccgaccatcgaattcttcgaacaacacagatactgttaagtggtgc

```

```

/codon_start=1
/protein_id="CAD12274.1"
/db_xref="GI:17048798"
/translation="MACTGWTSPPLVCLLLTGCFAEAGKLLVPMDSHWFTMRSV
VEKLILRHGVVVVPMPSVWOLGSLNCTVKTYSTVLTEDLDREFKAFARHAKQAV
RSTYSLMGSYNDIFDLFPNSCRSLFKDKKLYEYLKESSEDAVFLDPDNCGLIYAKY
FSLPSWVFARGILGHLYLEBGAQCPAPLSYFRILLGFSNMTFKERKPNRHLMHLEHL
LCHREFKALIEASEIILQTPVETDLYSHTSIMLLRTDFVLDYPRVPMNMFJGGIN
CHQGRFPMFEAEYINASGHEGIVFVSGMSVSEIPEKAMAIADALGKIPQTVLMRY
TGRFRPSLANANTILVWLPNODLLGHPMTAFITHAGSHGVYESICNGVPMVMPLEFG
DQMDNAKMETKAGVTLNLEYMTSEIDENALKAVINDKSYKENIMRLSLSHKDRPVE
PLDLAVFVWEVPMRHGAPHLRPAADHDLTWQYHSLDVIQGLLAVLVLTAPITFKCA
YGRKCLGKGRVKAHKSHT"
BASE COUNT      608 a  495 c  547 g      670 t
ORIGIN

```

```

Query Match      100.0%; Score 2320; DB 6; Length 2320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccagctgcagttctctgtggtggtcacagggtggaccagcccccctctctctatgtg 60
DB 1 CTCACGTCGAGTCTCTGATGCTGTGCACAGGCTGGACAGCCCCCTCTCTATGTGTG 60

QY 61 tctctgtctgcacctgtggtcttgcagagcgaggggaagctactgttagtgccatggat 120
DB 61 TGTCTGCTGTGACCTGTGGCTTTTCCCGAGCGCAGGAAAGCTACTGTAGTGCCCATGGAT 120

QY 121 gggagccactggttccacctgaggtcggtggtggagaaactcattctcaggggagataag 180
DB 121 GGGAGCCACTGGTTACCATGAGGTGCGTGGTGGAGAACTCATTTCTCAGGGGGCATAG 180

QY 181 ggggttagtcatgcagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
DB 181 GTGGTGTGTGATGTCAGAGGTGAGTGGCAACTGGGAAGATCACTGAAATTCACAGTG 240

QY 241 aagacttattcaacttataccctgaggtatcttgaccggaggttcaagggtctttgccc 300
DB 241 AAGACTTATTCAACTTATACCTTGAGGATCTTGACCGGGAGTTCAAGGCTTTTGCC 300

QY 301 catgctcaatggaaagcaagatcagagtatattctctattataaggggttcaataat 360
DB 301 CATGCTCAATGGAAAGCAGATGAGGATATATATTCCTATTAATGGGTTTCATCAAT 360

QY 361 gacattttgacttatttttttcaaatggcagaggttggtttaagacaaataatagta 420
DB 361 GACATTTTGTACTTATTTTTCATTAATGTCAGAGGTTTGTTTAAAGACAAAAATTAGTA 420

QY 421 gaataactaaagagaggtcttttgatgcagtggtttctctgatcccttttgataactgtgc 480
DB 421 GAATACTTAAGAGAGGTTCTTTTGTATGCAAGTGTTCCTGATTCCTTTTGATAACTGTGGC 480

QY 481 ttaa-tggtgccaaatatttctccctccctccgtggtcttcgcaggggaataactttgc 540
DB 481 TTAATGTGTGCCAAATATTTCTCCCTCCCTCCCTGCTGCTTCGCCAGGGGAATACTTTGC 540

QY 541 cactattctaaagaggtgcacagtgccctgtctctctctctctctctctctctctctctc 600
DB 541 CACTATTCTTGAAGAGGTGCACAGTGGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 600

QY 601 tttaggttctcagatgccactgactttcaaggagagagtagtcaggaaccacatcagtcacttg 660
DB 601 TTAGGGTTCTCAGATGCCATGACTTTCAAGAGAGAGATACGGAACCATCATCATGCACTTG 660

QY 661 gagguaatttattatgcacacgttttttcaaaaaatgccttagaataagctctctgaatt 720
DB 661 GAGGACATTTATTATGCCACCGTTTTTCAAAAAATGCCCTTAGAAATAGCCTCTGAAAT 720

QY 721 ctccnaacacctgttacgaggtatgatctctcagccacacatcaatttgggtgttgga 780
DB 721 CTCNAACACCTGTTACGAGGTATGATCTCTACAGCCACACATCAATTTGGTGTGGGA 780

```

```

QY 781 acggacttggtttggactatcccaaacccgtaatgcccaacatgatcttcatgtgtgt 840
DB 781 ACGGACTTTGTTTGGACTATCCCAAAACCCGCTGATGCCCAACATGATCTTCAATGTGGTGT 840

QY 841 atcaactgcactcagggagaaacggttgcttatggaatttgaagcctacataatgtctct 900
DB 841 ATCAACTGCCATCAGGGAAGCCGTGCTGCTATGGAATTTGAAGCCCTACATTAATGTCTCT 900

QY 901 ggagaaacatggaattgttggtttctcttgggtggtcgaatggtctcagaaaattccagagaag 960
DB 901 GGAGAAACATGGAATTTGTTGCTTTCTTTGGGATCAATGGTCTCAGAAATTCAGAGAGAAG 960

QY 961 aaagtatgcaaatgtctgacttgggggcaaaatccctcagaagaagtcctgtggcggtac 1020
DB 961 AAAGCTATGCAATTTGCTGCTTTGGGCAAAATCCCTCAGACAGTCCCTGTGGCGGTAC 1020

QY 1021 actggaacccgacccatcgaatcttgcgaacaacacagatacttggtaaagtggctaccacca 1080
DB 1021 ACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTTGTTAAGTGGCTACCCCAA 1080

QY 1081 aacgtctgtcttgctcaacccgatgacccgtgaccttattacccatgctggttcccaatggt 1140
DB 1081 AACGATCTGCTTGTGCTCCCGATGACCCGTGCTTATCACCATGCTGGTTCCTCCATGGT 1140

QY 1141 gtttatgaaagcatalgtcaatggcgttcccattggtggtgatgaccccttgggtggtgatcag 1200
DB 1141 GTTTTATGAAGCATATGCAATGGCGTTTCCCATGCTGATGATGCCCTTGTGTTGGTGTATCAG 1200

QY 1201 atggaacatcgaacgcctgagagactaaaggagctggagtgacctgaatgtctcggaa 1260
DB 1201 ATGGACAAATGCAAGCGCATGGAGACTAAGGAGCTGGAGTGACCTGGAATGTTCTGGAA 1260

QY 1261 atgacttctgaagatttagaaaatgctctaaaagcagtcataatgacaaaaagttacaag 1320
DB 1261 ATGACTTCTGAAGATTTAGAAAAATGCTCTAAAAGCAGTCAATCAATGACAAAAGTTACAAG 1320

QY 1321 gagaacatcagtcgacctccagcttcaacaggaccgcccggaggagcgcctgagacctg 1380
DB 1321 GAGAACATCATGGCCCTCTCAGCCTTCACAGGACCGCCGGTGGAGCGCCTGGACCTG 1380

QY 1381 gccgtgtctgggtgagtttggatgaggcaacaggcgccacacctgcgcccgca 1440
DB 1381 GCCGTGTCTGGGTGAGGTTTGTATGAGGCACAAAGGGCGGCCACACCTGCGCCCGCGCA 1440

QY 1441 gccacacacctccacctggtaccagtaacctcttgagcgtgattggttctcctctgccc 1500
DB 1441 GCCCACACCTCACCCTGTTACCACTACCATTCCTTGGACGTGATGGTTCCTCTTGGCC 1500

QY 1501 gtcgtgtgcagtgcccttaccacctttaaattgtgtgttattggtctacacggaatgc 1560
DB 1501 GTCGTGTGTCAGTGGCCCTTCATCACCCTTTAAATGTTGTGCTTATGCTACCGGAAATGC 1560

QY 1561 ttggggaaaaagggcgagttaaagaaagcccaaaatccaagacccattgagaagtgsgt 1620
DB 1561 TTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCAAAATCCAGAGACCCATTGGAGAGTGGGT 1620

QY 1621 gggaaaataaggtaaaatttgaaccattcccttagtcaatttcccaacttgaacacagaatc 1680
DB 1621 GGGAAATAAGGTAATAATTTGAACCATTCCTAGTCAATTCCTCAAACTTGAACACAGAAATC 1680

QY 1681 agtgttaattcattttcttatttaagaaacttctgataaataatcatcagccccag 1740
DB 1681 AGTGTAAAATTCATTTTATTTATTTAAGGAAATACCTTTGATTAATTAATATACGCCCCAG 1740

QY 1741 agtgctttaaataatctctctaaataaataatagactcgtagtcagtaagataatt 1800
DB 1741 AGTGCTTTAAAAAATCTCTTTAAAAATAAATAATAGACTCGCTAGTCAGTAAAGATATTT 1800

QY 1801 gaatatgacgtgccccctccggtctcttgaatcagatgaacatgacatgttttttcaaga 1860
DB 1801 GAAATATGTAACGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTTCAGA 1860

QY 1861 ggaagtgcagagcgctggcattcttagattacttcttacttgaaacatggcctggtt 1920

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 11:57:31 ; Search time 6248.67 Seconds
(without alignments)
5885.745 Million cell updates/sec

Title: US-09-835-082-3

Perfect score: 2320

Sequence: 1 ctacagtcagttctctgat.....ttctgtgtttctctacaact 2320

Scoring table: IDENTIFY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

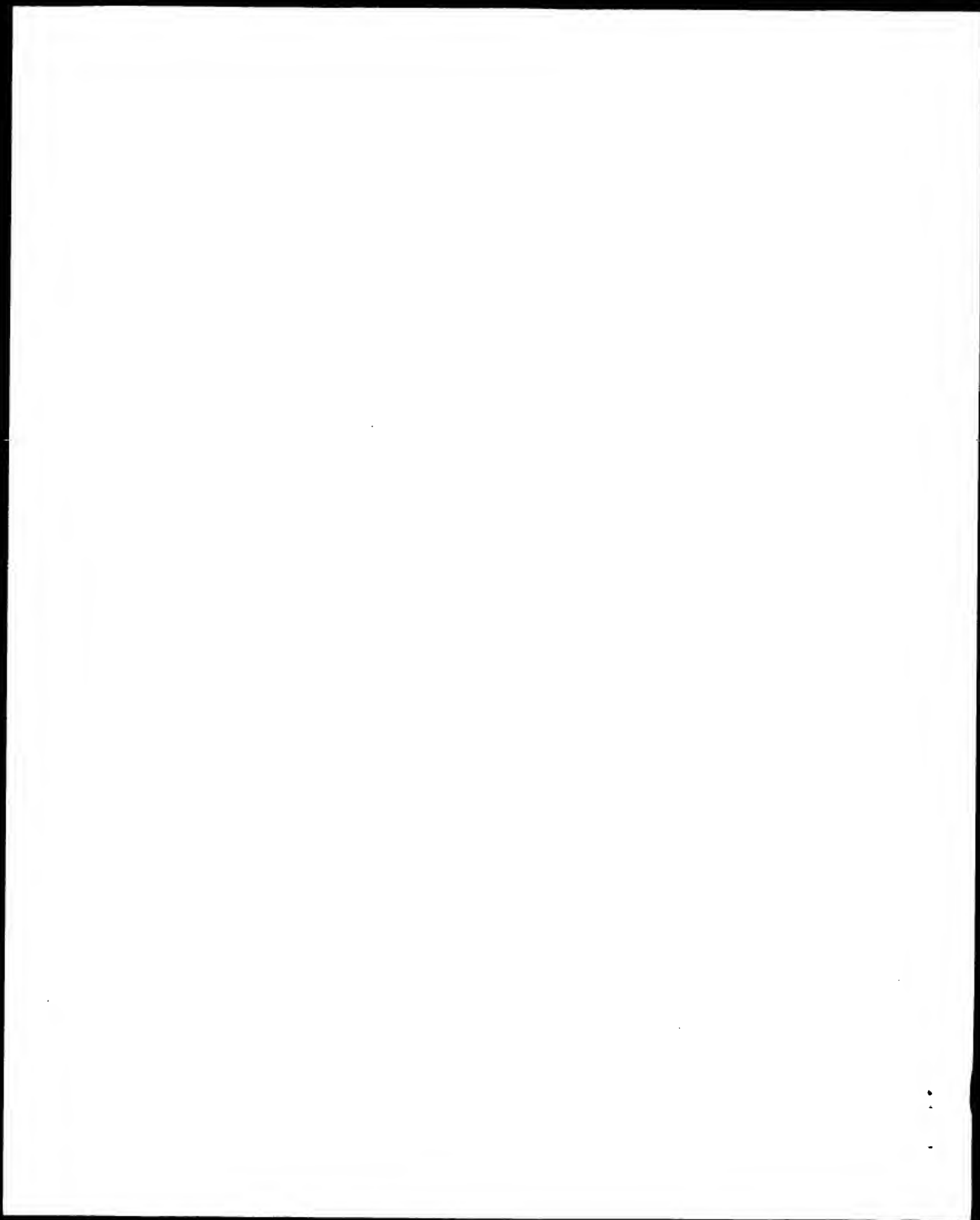
SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

1	2320	100.0	2320	6	AX286762	AX286762 Sequence
2	2320	100.0	2320	6	AX286764	AX286764 Sequence
3	2320	100.0	2320	9	AF056188	AF056188 Homo sapi
4	2218.4	95.6	2363	9	AF462267	AF462267 Homo sapi
5	2207.4	95.1	2394	9	AF462268	AF462268 Homo sapi
6	2194.6	94.6	2423	9	BC020971	BC020971 Homo sapi
7	2156.8	93.0	2392	9	S55985	S55985 Homo sapien
8	2076	89.5	2396	9	AF104336	AF104336 Macaca fa
9	1895.8	81.7	2003	9	AK023403	AK023403 Homo sapi
10	1636.6	70.5	2368	9	H08HUGBR2	M57951 Human billir
11	1629	70.2	2388	9	BC019861	BC019861 Homo sapi
12	1610.8	69.4	2351	9	H08HUGBR1	M57899 Human billir
13	1539.2	66.3	2422	6	AX336711	AX336711 Sequence
14	1539.2	66.3	2422	9	H08HUGTFA	J04093 Homo sapien
15	1493.8	64.4	1593	9	AF030310	AF030310 Homo sapi
16	1484.2	64.0	1593	9	HS089507	U89507 Human UDP-g
17	1477	63.7	1621	9	AF104338	AF104338 Macaca fa
18	1470.6	63.4	2440	9	AF104337	AF104337 Macaca fa
19	1463.6	63.1	1593	9	HS089508	U89508 Human UDP-g
20	1157.8	49.9	2241	10	AF461735	AF461735 Rattus no
21	1129.4	48.7	2203	10	BC019434	BC019434 Mus muscu
22	1123.4	48.4	1590	4	AF011902	AF011902 Oryctolag
23	1110.4	47.9	2155	10	MUSUGTP	L27122 Mus musculu
24	1105.2	47.6	2332	4	AB018478	AB018478 Ovis arie
25	1105	47.6	2282	10	AF461738	AF461738 Rattus no
26	1004.4	43.3	1561	9	H08HUGT03	M84124 Human UDP-g
27	1004.4	43.3	68770	9	AC006985	AC006985 Homo sapi
28	1004.4	43.3	176619	2	AC026497	AC026497 Homo sapi
29	1004.4	43.3	198872	9	AF297093	AF297093 Homo sapi
30	937	40.4	2301	10	RNU75903	U75903 Rattus norv
31	878.6	37.9	1255	9	AF297091	AF297091 Homo sapi
32	878.6	37.9	148438	9	AC019072	AC019072 Homo sapi
33	873	37.6	1602	9	AF360121	AF360121 Macaca mu
34	872.2	37.6	1647	9	AF104339	AF104339 Macaca fa
35	868.2	37.4	1602	9	AY029169	AY029169 Macaca mu
36	819.8	35.3	2816	4	AF039138	AF039138 Felis cat
37	799.8	34.5	183826	2	AC091778	AC091778 Papio cyn
38	794	34.2	2300	10	MMU16818	U16818 Mus musculu
39	788.4	34.0	2037	4	OCU09101	U09101 Oryctolagus
40	783.2	33.8	1765	10	RATUDPGTA	M34007 Rat bilirub
41	781.8	33.7	2343	10	AF461737	AF461737 Rattus no
42	777.9	33.6	1244	9	HS042604	U42604 Human UDP-g
43	777.8	33.5	1927	10	RATUDPGT	J02612 Rat UDP-glu
44	774	33.4	2239	10	RNU20551	U20551 Rattus norv
45	772.2	33.3	2484	10	RATUGT	J05132 Rat 3-methy

ALIGNMENTS

RESULT 1	AX286762	AX286762	Sequence 1 from Patent WO0180896.	2320 bp	DNA	linear	PAT 21-NOV-2001
LOCUS	AX286762	Sequence 1 from Patent WO0180896.					
DEFINITION	AX286762	Sequence 1 from Patent WO0180896.					
ACCESSION	AX286762	Sequence 1 from Patent WO0180896.					
VERSION	AX286762.1	GI:17048797					
KEYWORDS	human.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	1 (sites)						
AUTHORS	Ratain, M.J., Innocenti, F. and Iyer, L.						
TITLE	Flavopiridol drug combinations and methods with reduced side effects						
JOURNAL	Patent: WO 0180896-A 1 01-NOV-2001;						
FEATURES	ARCH DEVELOPMENT CORPORATION (US)						
source	Location/Qualifiers						
	1..2320						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	19..1611						
	/note="unnamed protein product"						
CDS							



Qy 1459 taccagtaccattccttgagctgattggtttcctcttgccgctgctgacagtggcc 1518
|||||
Db 1441 TACCAGTACCATTCTTGGAGCTGATTGGTTTCCTCTTGGCCGTCGTCGACAGTGGCC 1500
Qy 1519 ttcacaccttttaaatgttgcttatgctaccggaatgcttgggaaaaaaggcgca 1578
|||||
Db 1501 TTATCACCCTTTAAATGTGTGTGTATGGCTACCGGAATGCTTGGGGAAAAAGGCGCA 1560
Qy 1579 gttaagaaagcccaacaatccaagaccattga 1611
|||||
Db 1561 GTTAAGAAAGCCCAACAATCCAAGACCCATTGA 1593

Search completed: August 21, 2002, 11:57:31
Job time: 32834 sec

Db 239 AATACTACAAAGAAAAATCTATCCAGTCCGTATGACCAAGAGAGCTGAAGAACCGTT 358
QY 267 tcaagcttttgcacatgctcaatggaagcacaaglacgaagtatatattctattaa 346
Db 359 ACCAATCAATTTGGAAACAATCACTTTGCTGACGAGCATATTCCTAACTGCTCTCAGACAG 418
QY 347 tgggttcatacaatgaca-...ttttgacttatttttttcaaatgcagaggtttgttta 403
Db 419 AGTACAGGAATAACATGATTTTATTTGGCTGTACTTCATCAACATGCCAGAGCCTCCTGC 478
QY 404 aagaacaaaattatgaataactaaagagagttcttttgatgcagtggtttctogac 463
Db 479 AGGACAGGACACCTGAACTTCTTTAAGAGAGCAAGTTTGATCTCTTTTCACAGACC 538
QY 464 cttttgataactgtggttaattgttgccaaataatttccctccctccctccgtggtcttog 523
Db 539 CAGCTTACCCCTGTGGGTGATCCTGCTGAGTATTTGGGCCCTACCACATGCTGTGACCTCT 598
QY 524 ccagggaataactttggccaactatcttgaagaaggtgcacagtgccctgctctcttct 583
Db 599 TCAGGGGTTTTCGGTTCCTCGTGAGCATACATTTCAGCAGAAAGCCAGACCCCTGTGCTCT 658
QY 584 atgtccccaagaattctcttaggttctcagatgccaacttttcaagagagagtaaggga 643
Db 659 ACATTCCTCAGGTGCTACAAAGTTTTCAGACCAACATGACTTTTCCCAAGAGTGGCCA 718
QY 644 accacatcagcttgaggaacatttattatgcccacgttttttcaaaaatgcccotag 703
Db 719 ACTTCCCTTGAATTTGTTGGAGCCCTATCTATTATTTATGCTGTTTTCAAAAGTATGAGA 778
QY 704 aatagacctgaattctcoaaacacctgttacggagatgatctctacagccaacat 763
Db 779 AACTTCGATCAGCTGCTCTCAAGAGAGATGT---GGATATATCACTTTATCAGAGGTCT 835
QY 764 caattgtgttggagacgactttgttttggaactatcccaaacccgtgagtcaccaaca 823
Db 836 CTGTTTGGCTTTAAGATATGACTTTGTGCTTGAATATCTTAGCGGGTCATGCCCAACA 895
QY 824 tgaacttcattgttgatataactgccaatcaggaagcagcgttgcctatgggaattgaag 883
Db 895 TGGTCTTCATTTGAGGTATCACTGTATGATGATGATGATGATGATGATGATGATGATGAT 955
QY 884 cctacataatgtctcttgagaaacatggaattgtgtttctctctctctctctctctctct 943
Db 955 CCAATATATATGCTCTGGAGAACATGGAATTTGTGTTTCTCTCTTTGGGATCAATGTTCT 1015
QY 944 cagaattccagagagaaagctatgcaattgctgattgtgttggaataatccctcaga 1003
Db 1016 CAGAAATTCAGAGAGAAAGCTATGGCAATGCTGATGATGCTTTGGGCAAAACCCCTCAGA 1075
QY 1004 agtctctgtgggttacactggaaccgacccatcgaaatcttggaacaaacagatacttg 1063
Db 1076 CAGTCTCTGTGGGTGATCACTGGAAACCCGACCATCGAAATCTTGGCAACACAGTACTTTG 1135
QY 1064 ttaagtgtctaccacaaacagatctgttggtaaccgatacccgctgctttatcaacc 1123
Db 1136 TTAAGTGGCTACCCCAAAACAGATCTGCTTGGTACCCGATGATCCCGGCTTTATCACC 1195
QY 1124 atgtgttccctatgtgtttatgaagaacatgcaatggaatggaatggaatggaatggaat 1183
Db 1196 ATGTCTGTCTCCATGGTGTATGAAAGCATATGCAATGGCGTTTCCCATGGTGAATGATGC 1255
QY 1184 cctgtgtgtgatcagatggagaaatgcaagcgatggagactaaaggagctggagtgga 1243
Db 1256 CTTGTTTGTGTATCAGATGGCAATGCAAGGCGCATGGAGACTAAGGGAGCTGGAGTGA 1315
QY 1244 cctgtaattgttggaaatgactctctgaagatttgaataatgctctaaagcagtcata 1303
Db 1316 CCTCGAATGTCTGGAAATGACTTCTGAAGATTTAGAAATGCTCTAAAGCAGAGTCACTCA 1375
QY 1304 atgacaaaagtacaagagagaacatcatgscctctccagccttcacaagaccgcccg 1363
Db 1376 ATGACAAAAGTTCAAGGAGAACATCATGCGCTCTCCAGCCTTCACAGGACCCGCCG 1435

QY 1364 tggagcgtggaactggcgtgttctgtgggtggagtttggatgagggcacaagggcgcg 1423
Db 1436 TGGAGCCGTGGACCTGCCGTCTCTGGTGGAGTTTGTGATGAGGACAAAGGCGCGC 1495
QY 1424 cacactgcccccgagccccagcactcaactggtataceagttacacttcccttgagctga 1483
Db 1496 CACACTGCGCCCGCAGCCAGCCAGCTCACCTGTTACCAGTACCAATCTCTTGACCTGA 1555
QY 1484 ttgtttctctcttggccgtgctgacagtggtgcttcatcaccctttaaattgttgcct 1543
Db 1556 TTGTTTCTCTTGGCCGTCTGCTGACAGTGGCGCTTCATCACTTTAAATGTTGCTCTT 1615
QY 1544 atggtcacggaagtgtttgggaaaaagggcagtttaagaaagccacaaaatcccaaga 1603
Db 1616 ATGGCTACCCGAAATGTCTTGGGAAAAAGGCGAGTTAAGAAAGCCCAAAATCCCAAGA 1675
QY 1604 cccattgagaagtgggtgggaaaaaggtaaaattttgaaacctccctagctatttcca 1663
Db 1676 CCCATTGAGAGTGGGTGGGAAATAAGGTAAATTTTGAACCA-TCCTTAGTCAATTTCCA 1734
QY 1664 aactgaaaaacagaaatcagtggttaaatcatttattcttattaaaggaataacttgcac 1723
Db 1735 AACTTGAANAACAGATCAGTGTAAATTCATTTTATTCTTATTAAAGAAATACCTTGCAAT 1794
QY 1724 aaattaatcagccccagagtg-ctttaaaaaatctctttaaataaaaaataatagactcgc 1782
Db 1795 AAATTAATCAGCCCCAGAGTGTCTTTAAAAAATTCCTTTAAATAAATAATAGACTGCG 1854
QY 1783 tagtcagtaagaatatttgaatgatgtgctgccccctccgggtgtcttttgatcaggaatga 1842
Db 1855 TAGTCAGTAAGATATTTTGAATATGATCTGTCGCCCTCCGGAGTCTTTGATCAGGATGA 1914
QY 1843 catgtccatttttcagagagcgtgcagacagctggtcattctagattactttcttaact 1902
Db 1915 CATGTGCCATTTTTCAGAGACGTGCAGACAGCTGGCATTCTAGATTACTTTCTTACT 1974
QY 1903 ctgaacatggcctgtttgggagtgcggaattcgaagtggtcccccagcgtgccccact 1962
Db 1975 CTGAACATGGCTGTGTTGGGAGTGGGGATTCAAAAGTGGTCCACAGCTGCCCTACT 2034
QY 1963 gcaattggcagttttaaacttcttctgtgtcttcagagctggttcagattgactcttaa 2022
Db 2035 GCAATGGCAGTTTAACTTATCTTTTGGCTTCGACAGATGGTTCG-ATTGATCCITAA 2093
QY 2023 ccaataatggtcagtcctcatctctgctg-cttcatagggtgccacctgtgtgtttaa 2081
Db 2094 CCAATAATGGTCAGTCTCATCTCTGCTGACTTCATAGGTGCCACCTGTG-GTTTAA 2152
QY 2082 agaagggaagcttttacccttagaagtagtggaatgaatgaatggcttgagtgac 2141
Db 2153 AGAAGGAGAGCTTTGTACTTTAGTGTAGGTGAATGAATGAATGAATGGCTTGGAGTGCAC 2212
QY 2142 tgaaacagcatatgatttcttcttgggaaaaaagaatgatgatgaatgtgtggg 2201
Db 2213 TGAGAACAGCATATGATTTCTGTTTGGGAAAAAGAAATGATGCTATGAAATGGTGGG 2272
QY 2202 tgggtatttgaagagataaactctgcttctatgtcaaatggagctgaattgtataaaacc 2261
Db 2273 TGGTGTATTGAGAGATAAATCATTCCTTATGTCAAAATGGAGCTGAATTTGATTAANACC 2332
QY 2262 caaaatcacgctatgaagtgctgggaaagtttac-tttttcttgatgtttctcacaact 2320
Db 2333 CAAAAATACAGCTATGAAGTGTGGGCAAGTTTACTTTTCTGTGATGTTTCTCTACAAC 2392

RESULT 15

AF030310
LOCUS AF030310 1593 bp mRNA linear PRI 26-OCT-2000
DEFINITION Homo sapiens UDP-glucuronosyltransferase 1A8 (UGT1A8) mRNA,
complete cds.
ACCESSION AF030310
VERSION AF030310.1 GI:2613043

Db 1256 CCTTGTGGTATCAGATGGACAAATGCAAAAGCGCATGGAGACTGAAGGAGCTGGAGTGA 1315
 Qy 1244 ccclgaatttctggaatgaactctgaagatttagaaatttagaaatgctctaaagcagctatca 1303
 Db 1316 CCCTGAATCTTCTGGAATGACCTTCTGAAGATTTAGAAATGCTCTAAGACATCATCA 1375
 Qy 1304 atgcacaaagttaacaaaggaacacatcatcgccctctccagccttccacaggaaccgcccgg 1363
 Db 1376 ATGACAAAGTTACAAAGGAACATCATGCGCCTCTCCAGCTTCACAGGACCGCCCGG 1435
 Qy 1364 tggagccgctggaacctgcccgtgttctggtggtggtggtggtggtggtggtggtggtggtggt 1423
 Db 1436 TGGAGCCGCTGGACCTGGCCGTCTTCTGGGTGGAGTTGTGTGATGAGGCACAGGGCGCG 1495
 Qy 1424 cacactcgcccgccgaagcccaagcactcaactggtaccagatcattctcttggagctga 1483
 Db 1496 CACACCTGGGCCCGCCGAGCCAGCCCTCACCTGGTACCACTTACCACTTCTCTGGAGCTGA 1555
 Qy 1484 ttggttctcttggccgtgtgctgacagtggtggtggtggtggtggtggtggtggtggtggt 1543
 Db 1556 TTGGTTCTCTTCTGGCCGTGCTGCTGACAGTGGCTTCTACCTTTAAATGTTGTCCTT 1615
 Qy 1544 atggtaccgaaatgcttggggaaataaggtgaaataaggtgaaataaggtgaaataaggtgaa 1603
 Db 1616 ATGGCTACCCGAAATGCTGGGGAAAGGAGGAGTGAAGAAAGCCACAAATCCCAAGA 1675
 Qy 1604 cccattgagaagtggtggtggaataaggtgaaataaggtgaaataaggtgaaataaggtgaa 1663
 Db 1676 CCATTTGAGAGTGGGTGGGAAATAGGTAAATTTTGAACCA-TCCCTAGTCATTTTCCA 1734
 Qy 1664 aacttgaacacagaatcagtgtaattcatcttattcttattcttattcttattcttattcttattct 1723
 Db 1735 ACTTGAACACAGAAATCAGTGTGTAATTTCAATTTATTTATTTATTTATTTATTTATTTATTT 1794
 Qy 1724 aaattaatcagcccgagtg-cctttaaataatcttcttaaataatcttcttaaataatcagctgc 1782
 Db 1795 AAATTAATCAGCCCGCAGTGTCTTTAAATAATCTCTTAATAATAATAATAATAATAATAATA 1854
 Qy 1783 tagtcagtaagatatttaataatgattatgctgcccctccggtgctttagatcagatga 1842
 Db 1855 TAGTCAGTAAGATATTGAATATGATATGATGTCGCCCTCCGAGTCTTTGATCAGATGA 1914
 Qy 1843 catgtgcaattttcagaggaacgctgagacagagctggcattctagattcttcttact 1902
 Db 1915 CATGTGCATTTTTCAGAGAGAGCTGACAGACAGCTGGCATTTAGATTTACTTTTCTTACT 1974
 Qy 1903 ctgaacatggtcctgttgggtggtggtggttcaaaagtggtccacgcgtgcccctact 1962
 Db 1975 CTGAACATGGCCTGTTTGGGAGTGGGGATTCAAAGTGTGCCACGGCTGCCCTTACT 2034
 Qy 1963 gcaaatggcagtttaattcttattcttcttcttcttcttcttcttcttcttcttcttcttct 2022
 Db 2035 GCAAAATGGCAGTTTAACTTATCTTTTGGCTTCTGCAGATGTTGTC-ATTGATCTCTTAA 2093
 Qy 2023 ccaataatggtcagctcctatctctgctctg-cctcataggtgcccacttctgctgtttaa 2081
 Db 2094 CCAATAATGTCAGTCTCATCTGCTGCTGACTTCTATAGTGGCCACTTGTG-ATTGTA 2152
 Qy 2082 aagaaggaagctttgtaacatttagagtgtaggtgaaatgaatgaatggtggtggtggtggtggt 2141
 Db 2153 AGAAGGGAAGCTTTGACCTTTAGAGTGTAGGTGAAATGAATGAATGCTTGGAGTGCAC 2212
 Qy 2142 tgagaaacagcatatgatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2201
 Db 2213 TGAGAACAGCATATGATTTCTTGTCTTGGGGAAAAAGAAATGATGCTATGAATTTGTTGGG 2272
 Qy 2202 tgggtatttgagaagaataatcattgcttcttcttcttcttcttcttcttcttcttcttcttct 2261
 Db 2273 TGGTGTATTTGAGAAGATTAATCATTTGCTTATGCAAAATGGAGCTGAATTTGATAAAACC 2332
 Qy 2262 caaatacagctatgagtgctggggcaagtttac-ttttttctgattgttctctcaact 2320
 Db 2333 CAAAATACAGTATGAAGTGTGGGCAAGTTTACTTTTTTTTCTTGATGTTTCTTCTACACT 2392

RESULT 14
 HUMUGT1FA
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

gene

CDS

polyA_signal

polyA_signal

polyA_site

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 1870; Cnserve

0; Mismatches

398; Indels

12; Gaps

8;

Qy 47

Db 119

Qy 107

Db 179

Qy 167

Db 239

Qy 227

166

178

238

226

298

286

106

178

238

226

298

286

106

178

238

226

298

286

106

178

238

226

298

286

HUMUGT1FA
 Homo sapiens phenol UDP-glucuronosyltransferase (UDPGT) mRNA,
 complete cds.

J04093
 J04093.1 GI:624724

UDP-glucuronosyltransferase; phenol.
 Homo sapiens kidney cDNA to mRNA.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Rutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2422)

Harding,D., Fournel-Gigleux,S., Jackson,M.R. and Burchell,B.
 Cloning and substrate specificity of a human phenol

UDP-glucuronosyltransferase expressed in COS-7 cells
 Proc. Natl. Acad. Sci. U.S.A. 85 (22), 8381-8385 (1988)

89042187
 On Jan 18, 1995 this sequence version replaced gi:340517.

Location/Qualifiers
 1..2422

/organism="Homo sapiens"
 /db_xref="taxon:9606"

/tissue_type="kidney"
 1..2422

/gene="UDPGT"
 88..1683

/EC_number="2.4.1.17"
 /codon_start=1

/product="UDP-glucuronosyltransferase"
 /protein_id="AA61251.1"

/db_xref="GI:624725"
 /translation="MACLLRSQRIAGVFFELALMGVVDKLLVVPDGSIMLSMDK

IEVLSADRGHEIVVYVPEVNLULLKRYTRDIYPIYDOELKRYISQFNNHFAER
 SEFAPQTEYRNNMIVIGLYFINCSLLQDRDTLNFKESKFDALTDPAIPCCGVILA

EYLGPSVYLEFGPFCSELTFRSPDPVSIPRCYTKFSDHMTFSQVNFVNLIE
 PLYFLCLSKYELASALRDVDITLSEVSWLLRYDFVLEPRVPMNMFVIGI

NCKRKDLDSQFEAYINAGCEHIVFSLGSMVSEIPEKKAMAIADALGKNPQTVLWR
 YTKTRPSNLANTILVKVLPONDLLGHPMTAFITHAGSHGVYESICNGVPMVMPLF

GDQMDNARMETKAGVTNLNLEMTSELENALKAVINDKSKYKINIMLSLHKDRPV
 EPDLIAVFWFEVVRHKGAPHLRPAADHDLTWYQVHSLDVLGFLAVLVTAFITFKCC

PVGYPKCLGKGGKVRKHKSKTH"
 2396..2401

/gene="UDPGT"
 2404..2409

/gene="UDPGT"
 2422

/gene="UDPGT"
 653 a 516 c 555 g 698 t

BASE COUNT
 ORIGIN

Query Match
 Best Local Similarity
 Matches 1870; Cnserve
 0; Mismatches
 398; Indels
 12; Gaps
 8;

Qy 47

Db 119

Qy 107

Db 179

Qy 167

Db 239

Qy 227

166

178

238

226

298

286

106

Qy	1805	atgtatcgtgccccctccggtgtcttttgatcaggtgacatgtgccatttttcagaggac	1864			
Db	1811	ATGTATCGTCCCGCTCCGCTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC	1870			
Qy	1865	gtcagacagcgtggcattctcagattacttttctactctgaaacatggccctgtttggga	1924			
Db	1871	GTCCAGACAGCGTGGCACTCTAGATTTACTTTCTTACTCTGAAACATGGCCCTGTTGGCA	1930			
Qy	1925	gtcgggattccaagggtgccaccgcgtgccctactgcaaatggccgttttaacttta	1984			
Db	1931	GTCCGGGATTCAAAGTGGTCCCAACCGCTGCCCTTACTGCCAAATGGCAGTTTAAATCTTA	1990			
Qy	1985	tcttttggtcttcgcagatgggttccaatgatccttaaccaaataatggctcagtcctcattc	2044			
Db	1991	TCTTTTGGCTCTGCAGATGGTGTCAATTGATCCTTAACCAATAATGGTTCAGTCCCTCATC	2050			
Qy	2045	tctgtcctgttcaatagggtgccacctgtgtgtttaaagaaggaagcttctgataccttta	2104			
Db	2051	TCGTCTGCTCTCATAGGTGGCCACCTGTGTGTGTTTAAAGAAGGAAGCTTGTACCTTTA	2110			
Qy	2105	gagtgtagtgaaatgaatgaatggcttgagtgcactgagacagcatatgattctctg	2164			
Db	2111	GAGTGTAGGTGAATGAATGAATGGCTTGGAGTGCCTAGAGACAGCATATGATTTCTTG	2170			
Qy	2165	ctttggaaaaaagaatgaactatgaaattggtgggtgtgtatttgagaagaataatca	2224			
Db	2171	CTTTGGGAAAAAGATGATGCTATGAAATGGTGGGTGTGTAITTTGAGAGATTAATCA	2230			
Qy	2225	tgtcttatgtccaaatggagctgaatttgataaaaaaccccaaaatacagctatgaagtgtg	2284			
Db	2231	TTGCTTATGTCAAAATGGAGCTGAATTTGATPAAAAACCCAAAATACAGCTATGAAGTCTG	2290			
Qy	2285	ggcaagtttaacttttttctgtatgtttcttcacaact	2320			
Db	2291	GGCAAGTTTACTTTTTTCTGATGTTCCTTACAACCT	2326			
RESULT 13						
AX336711						
LOCUS	AX336711	2422 bp	DNA			
DEFINITION	Sequence 7220 from Patent WO0194629.					
ACCESSION	AX336711					
VERSION	AX336711.1					
KEYWORDS	GI:18127430					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.					
1 (sites)						
Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,						
Horriggan, S., Soppet, D.R. and Weaver, Z.						
Cancer gene determination and therapeutic screening using signature						
gene sets						
JOURNAL	Patent: WO 0194629-A 7220 13-DEC-2001;					
Avalon Pharmaceuticals (US)						
Location/Qualifiers						
1..2422						
/organism="Homo sapiens"						
/db_xref="Taxon:9606"						
BASE COUNT	653	a	516 c 555 g 698 t			
ORIGIN						
Query Match 66.3%; Score 1539.2; DB 6; Length 2422;						
Best Local Similarity 82.0%; Pred. No. 0;						
Matches 1870; Conservative 0; Mismatches 398; Indels 12; Gaps 8;						
Qy	47	tctctctatgtgtgtctgtctgtcgtacactgtgcttgcgcaggcagggaagctactg	106			
Db	119	TTTCTCGACGGGTTTCTCTTCTTAGCTCTTTTGGGGCATGSGTTGTAGTGTCAAGACTGCTG	178			
Qy	107	taqtgcacctgaatggagacacatgattccaacctgaatgaatcgtgtgtgtgaagaactcattc	166			

```

/protein_id="AAA63195.1"
/db_xref="GI:184473"
/translation="NAVESQGRPLVLGLLCLVLPVVSAGKILLIPYDGSNLSML
CAIQLOQRHEITVVLADSLYIRDAFVTLKTYPPROREDVKSFPVSLGHNLFN
DSFLQVITKTKYIKKDSAMLLSCSHLLNKLMALESFSDVMTDPFLPCSPIV
AOYLSLPTVFLLHLPCLSPFATQCPNPSYVPRPLSSSDHMTFLQRYKMLLIAS
ONFLCDVVSYPATLASPELQREVTVDLLSSASWLFKRSDFVDPRIPIPNWVFG
GLNCLQNELSDPEAYINASGEHIVFSLGSVSEIPEKKAWADALGRIQTVL
WRTYGRNPLANNITLVKLPONDLLGHPMTFRFTHAGSHGYESIICNGVPVWMP
LPEQDMNAKMETKGAGVTLNLEMTSEDLLENALKAVINDKSYKENINRLSSLHKD
VFPLDLAYVWFYFMRHKGAPHLRPAADHLWYQYHSLDVIGVLLAVLTVAFTEK
CCAYGKCKLGRKGRYKKAHKSHT"
43.75
sig_pept.de
/genes="UGT1"
/notes="G00-120-007"
76..1614
mat_pept.de
/standard_name="bilirubin UDP-glucuronosyltransferase
isozyme 1"
/EC_number="2.4.1.17"
/product="UDP-glucuronosyltransferase 1"
BASE COUNT 601 a 540 c 557 g 653 t
ORIGIN

Query Match 69.4%; Score 1610.8; DB 9; Length 2351;
Best Local Similarity 82.7%; Pred. No. 0;
Matches 1866; Conservative 0; Mismatches 387; Indels 3; Gaps 2;

Qy 68 tgcgtacctgtggttgcgagcgaggaaactaactgtagtgcgcatgatggagcc 127
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 TGTGCGGCCGACGTGTCCTCCATGCTGGGAAGATACACTGTATGCCATCCATGGCAGCC 130
Qy 128 actggttcaaccatgagtcogtgggtggggaactcaattctcaggggggcatgaggtgtg 187
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
131 ACTGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAATAGTTG 190
Qy 188 tagtcatgccagaggtgagttggcaactgggaagatcactgaatgacagtgacagtgaaactt 247
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 TCCTAGCACCCTGACGCTGCTGTATACACAGACGAGGCATTTTACACCTTGACACGCT 250
Qy 248 attcaacttataacccctggagagatctggaccggagtt--caaggctttggccatgc 305
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 ACCCTGTCACATCCAAAGGGAGGATGTGAAAGAGTCTTTTGTAGTCTCGGCGCATATG 310
Qy 306 tcaatggaaagcaagaagtagatataattctctatitaaatgggttccatacaatgacat 365
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
311 TTTTGTGAGATGATTTCTTCGACGCTGTGATCAAAACATACAAAGAAATAAAGG 370
Qy 366 ttttgact-tattttttcaaatgacaggagttgttttcaagacaaaaattagtagaat 424
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 ACTCTGCTATCTTTGCTGTGCTGCTCCACCTTACTGCACAAAGAGGCTCATGCGCT 430
Qy 425 acttaaggagaggttttttgatgagtggtttctcgatccttttgataactgtggttaa 484
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 CCTCGGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCTTTCTCTTCCCTGACGCCCA 490
Qy 485 ttgttgcacaaatattctccctccctccctgggttcttcgacggggaatactttgcaat 544
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
491 TGTGCGGCCAGTACTGCTCTCTGCGCCACTGTATTTCTTGTGATGCACTGCGCATGCGCC 550
Qy 545 attcttgaagaaggtgcacagtgccctgctcctctttctctatgccccagaattctcttag 604
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 TGGAAATTTGAGGCTTACCGAGTCCGCCCAACCATTCCTACTGCTCCCGCCGCTCTCTCT 610
Qy 605 ggttctcagatgcatgactttcaaggagagagtagcggaccacatcatgacattcgtgagg 664
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 CTCATTCAGATCAGTACGCTTCTGTCAGCGGGTGAAGAACATGCTCATTTGCTCTTTCAC 670
Qy 665 aacatttattatgccacggttttttcaaaaaatgccttagaaaaatagcctctgaaattctcc 724
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 671 AGAACTTTCTGTGGCAGCTGTTTATTTCCTCCGTATGCAACCCCTTGCTCAGAAATTCCTTC 730
Qy 725 aaacactgttacgagtagatgactctctacagccacacacatcaatttggttggtgcaacgg 784
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
731 AGACAGAGTGACTCTCCAGACCTATTGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790
Qy 785 actttgtttgactatccccaaacccgtgtagtgcacacacacacacacacacacacacacac 844
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
791 ACTTTGTGAAGGATTACCTAGGCCCATCATGCCCAATATGCTTTTGTGTTGTTGTTGTTGTT 850
Qy 845 actgcacatcagggaacccgttgcctatggaattgaagcctacacacacacacacacacacac 904
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
851 ACTGCCCTTACCANAATCCACTATCCAGGAATTTGAAGCTATACATTAATGCTTCTCTGAG 910
Qy 905 aacatggaattgtgtgttttctcttcttgggacacacacacacacacacacacacacacacac 964
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
911 AACATGGAATTTGTTTCTCTTTGGGATCAATGGTCTCAGAAATTCAGAGAAGAAAG 970
Qy 965 ctatggcaattgctgtagtcttggcacaacacacacacacacacacacacacacacacacacac 1024
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
971 CTATGGCAATTTGCTGATGCTTTGGGCAAAATCCCTCAGACAGCTCTGTGTTGGGGTACACTG 1030
Qy 1025 gaaccgcacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1084
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1031 GAACCCGACCATCGAATCTTTGGCAACACACAGACTTGTTAAGTGGCTAGCCCAAAACG 1090
Qy 1085 atctgctgtgtccacccgatgacccgtgctcttataccacacacacacacacacacacacacac 1144
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1091 ATCTGCTTSGTCAACCGCATGACGCGCTGCTTTATCACCCATGCTGTTGCCATGCTGTTT 1150
Qy 1145 atgaagacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1204
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1151 ATGAAAGCATATGCAATGGCGTTGCCATGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1210
Qy 1205 acaatgcaagcgcatgagacactaaggagctggagtgacccctgaatgtcttggaatga 1264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1211 ACAATGCAAGCGCATGAGACTAAGGAGCTGAGAGTACCCCTGAAATGCTTCTGGAATGA 1270
Qy 1265 ctcttgaagatttagaaaaatgctctaaagcagtcacacacacacacacacacacacacacacacac 1324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1271 CTCTGAAAGATTAGAAATATGCTCTAAAGCAGTCAATGACAAAAGTTTACAAAGGAGA 1330
Qy 1325 acatcatgcctctccagcttccacagcagcccggtggagccgctggaacctggaacctggccg 1384
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1331 ACATCATCGCTCTCCAGCTTTCACAGAGACGCCCGGTGGAGCGCTGAGCTGGCCG 1390
Qy 1385 ttttctgtgtgagattgttga tgaagcaagagggcgcgacacacacacacacacacacacacacac 1444
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1391 TGTCTGTGTGGAGTTTGTATGAGGCAACAGGGCGGCCACACCTGCGCGCCGCGACGCC 1450
Qy 1445 acgacctcaccctgggtacacagtagtaccattctcttggagctgattggtttctcttggcctg 1504
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1451 AGGACCTCACCTGTTACAGTACCATTCCTTGGAGCTGATTGTTTCTCTTGGCGCTG 1510
Qy 1505 tgcacacatgctctcatcacctttaaattgttcttattgctacacacacacacacacacacacacac 1564
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1511 TGCTGACAGTGGCTTTCATCACCCTTAAATGTTGTCTTATGGCTTACCGGAAATGCTG 1570
Qy 1565 ggaataaaacggcgagttgaagaagcccaaaaacacacacacacacacacacacacacacacacacac 1624
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1571 GGAAAAAGGGCGAGTTAAGAAAGCCCAAAATCCAAAGACCCATTGAGAAAGTGGGTGGGA 1630
Qy 1625 aataaggttaaaatttgaacacacacacacacacacacacacacacacacacacacacacacacacac 1684
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1631 AATAAGGTPAAATTTTGAACCAATTCCTTCTAGTCAATTTCCAAACCTTGAAGAAACAGAAATCAGTG 1690
Qy 1685 ttaatactatttattcttatttaaggaaaacttttgataaaataacagccccacagtg 1744
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1691 TTAATAATTCATTTATTTCTTTATTAAGGAAATACTTTGCATAAATTAATTCAGCCCCCAGAGTG 1750
Qy 1745 cttttaaaaaattctcttaataaaaaataatagactgtagtcaagtaaaatatttgaat 1804
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1751 CTTTAAAAATTTCTTTAAATAAAAAATAATAGACTCGCTAGTAAAGATATTGAAT 1810

```

```

Db 726 AACTCGCATCAGCTGCTCCTCAAGAGAGATCTGGATATATATCACTTTATATCAGAAGTCT 785
QY 764 caatttggttggtgcaaacggaacttggttggtgactatcccaaacccggtgagtgcccaaca 823
Db 786 CTGTTTGGCTGTGTAGATATGACTTGTCTTGAATATCCTTAGCCGGGTTCATGCCCAACA 845
QY 824 tgattcttcatttggttggtatcaactgcatcagggaagccggttgccctatggaattgaaag 883
Db 846 TGGTCTTCATTTGGAGTATCACTGTAAGAAGAGAGGAAGACTTGTCTCAGGAATTTGAAG 905
QY 884 cctacataatctctctggaacacatggaattggtgtttctcttcttggaatacaatggtct 943
Db 906 CTAATTAATGCTTCTGGAGACATGGAAATGTGTGTTTCTTCTTGGGATCAATGGTCT 965
QY 944 cagaataatccagagaagaaactatgcaattgctgattgcttggtgggcaaaatccctcaga 1003
Db 966 CAGAAATCCAGAGAAGAAAGCTATGGCAACTGCTGATGCTTTGGGCAAAATCCCTCAGA 1025
QY 1004 cagtcctgtggtggtacactggaacccgaccatcgaattcttgcaaacacacgatacttg 1063
Db 1026 CAGTCTGTGGCGGTACTCTGGAACCCGACCATCGAATCTTGGCAACAACACGATCTTG 1085
QY 1064 ttaagtggctaccacaaacacgactgcttggtcaccgatgacccctgacctttatcaacc 1123
Db 1086 TTAAGTGGCTTACCCCAAAACGATCTGTTGGTCACCCGATGACCCGCTGCTTTATCACCC 1145
QY 1124 atgctgtgtcccatggtttatgaaagcatatgcaatggcggtcccatggtgtagatgc 1183
Db 1146 ATGCTGTTCCCATGCTGTTTATGAAGCATATGCAATGGGCTTCCCATGTTGATGATGC 1205
QY 1184 ccttggttggtgacagatgagacaatgcaaatgcaaacgcatgagagactaaggagctgagtg 1243
Db 1206 CTTGTTTGGTGATCAGATGAGCAATGCAAGCGCATGGAGACTTAAGGAGCTGGAGTGA 1265
QY 1244 cccgtaattgtctgaaatgactctgaagatttagaaatgctctaaagcagtcac 1303
Db 1266 CCTGTAATGTTCTGGAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCA 1325
QY 1304 atgcaaaaagtataaagagagaacatcatcgccctctccagcttcacaaaggaccgccgg 1363
Db 1326 ATGCAAAAATTTACAAAGGAACATCATGCGCCTCTCCAGGCTTTACAAAGGACCCTCCGG 1385
QY 1364 ttagcgcgttggaactggccgttgcttggtggtgagttgtagaggaacaaaggcgccg 1423
Db 1386 TGGAGCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTGTCTATGAGCAACAAGGCGCGC 1445
QY 1424 cacacctgcccccgagcccaacacacacacacacacacacacacacacacacacacacac 1483
Db 1446 CACACCTGCGCCCCGAGCCGACACCTCACCTGGTACAGTACCATTCCTTGGACGTGA 1505
QY 1484 ttggtttcctcttgccgtcgctgctgacagtggccttcacaccttttaaatgttgctt 1543
Db 1506 TTGTTTCTCTTGGCGCTGCTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTT 1565
QY 1544 atggtaccggaatgcttggtgggaaaaaaggcgagttgaagaaagcccaataccaaga 1603
Db 1566 ATGGCTACCGGAAATGCTTGGGGAAGAAAGGGGAGTTAAGAAAGCCCAATCCAAGA 1625
QY 1604 cccattgagaagtgggtgggaaaaaaggtaaaatttgaaccattccctagtcattcca 1663
Db 1626 CCAATTGAGAAGTGGGTGGGAATAAGTAAATTTTGAACCATTCCTAGTCAATTTCCA 1685
QY 1664 aacttgaaacacaaatcaagttaaatccattttcttatttaagaaataactttgcat 1723
Db 1686 AACTTGAACACGAATCAGTGTAAATTCATTTATCTTATCTTATTAAGGAATACTTTGCAT 1745
QY 1724 aaattaatcagccccagagtgctttaaaaaattctttaataaaaaataatagactcgct 1783
Db 1746 AAAATTAATCAGCCCCAGAGTGTCTTAAAAAATTTCTTAAAAAATAAATAGACTCGCT 1805
QY 1784 aqtcagtaagatattgaatgtagtgcgtgccccctcggtgcttcttgatcagatgac 1843

```

```

Db 1806 ACTCAGTAAGAATATTTGAATATGTATGTCGCCCTCCGGTCTCTTTGATCAGGATGAC 1865
QY 1844 atgtgccatttttcagaggagcgtgcagacagcgtggcattcttagattacttttttactc 1903
Db 1866 ATGTGCCATTTTTCAGAGGACGTCGACAGACAGGCTGGCATTCTAGATTACTTTTCTACTC 1925
QY 1904 tgaacatgacctgtttggagtgccggaatcaaaagtgatcccaaccctgccccctactg 1963
Db 1926 TGAACATGGCTGTGTTGGGAGTGGCGGATTCAAAGGTGGTCCACCGCTGCCCTACTG 1985
QY 1964 caaatggcagtttaattctatctcttttggcttctgcagatggttgcaattgatccttaac 2023
Db 1986 CAAATGGCAGTTTAACTTATCTTTTGGCTTCTGTCAGATGGTTGCAATGATGATCCTAAC 2045
QY 2024 caataatgctagtcctcatctctctctctctctctctctctctctctctctctctctct 2083
Db 2046 CAATAATGGTCAGTCCTCATCTCTGCTGCTTCATAGGTGGTCCACCGCTTGTGTGTTAAAG 2105
QY 2084 aagggaagctttgtaccttttagagtgtagtggaatgaatgaatgaatggagtgccactg 2143
Db 2106 AAGGGAAGCTTTGTACCTTTAGAGTGTAGTGAATGAATGAATGGCTTGGAGTGCACCTG 2165
QY 2144 agaacgcataatgattctctctctctctctctctctctctctctctctctctctctctct 2203
Db 2166 AGAACAGCATATGATTTCTGCTTTGGGAAAAGAAATGATGCTATGAATTTGGTGGTG 2225
QY 2204 gtgtattgagaagaataatcatcttctctctctctctctctctctctctctctctctct 2263
Db 2226 GTGTATTTCAGAAAGATAATCATTTGCTTATGTCAAATGGAGTGAATTTGATAAAACCCA 2285
QY 2264 aaatacagatatgaagtgtggaagtttctctctctctctctctctctctctctctctct 2320
Db 2286 AATACAGCTATGAAGTCTGCGGCAAGTTACTTTTTTCTGATGTTTCTCTACAACT 2342

RESULT 12
LOCUS HUMHUGB1
DEFINITION Human bilirubin UDP-glucuronosyltransferase isozyme 1 mRNA.
ACCESSION M57899
VERSION M57899.1 GI:184472
KEYWORDS UDP-glucuronosyltransferase; bilirubin UDP-glucuronosyltransferase.
SOURCE Human adult female liver, cDNA to mRNA, clones Z6, Z11 and Z6MB2.
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2351)
AUTHORS Ritter,J.K., Crawford,J.M. and Owens,I.S.
TITLE Cloning of two human liver bilirubin UDP-glucuronosyltransferase
cDNAs with expression in COS-1 cells
JOURNAL J. Biol. Chem. 266 (2), 1043-1047 (1991)
MEDLINE 91093210
FEATURES
Location/Qualifiers
1..2351
/organism="Homo sapiens"
/isolate="AK"
/db_xref="taxon:9606"
/map="2"
/clone="Z6"
/sex="female"
/tissue_type="liver"
/dev_stage="adult"
/tissue_lib="lambda-ZAP"
1..2351
/gene="UGT1"
16..1617
/gene="UGT1"
/standard_name="bilirubin UDP-glucuronosyltransferase
isozyme 1"
/codon_start=1
/product="UDP-glucuronosyltransferase 1"

```


QY 846 ctgcatcagggaagcgcttgctatggaattgaagcctacattaatgcttcttgaga 905
 Db 481 CTGTCATCAGGGAAGCGCATGCGCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAGA 540
 QY 906 acatggaattggtgttcttcttgatcaaatggtctcagaataatcagaagaagaagc 955
 Db 541 ACA:GGAATTTGGTGTCTCTTTGGGATCAATGCTCTCAGAAATTCAGAGAAGAAGC 600
 QY 966 tatggaattgctgatgcttcttgaggcaaatccctcagacagctcgtgctggtacactgg 1025
 Db 601 TATGGCAATTTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCTCTGTGGCGGTACACTGG 660
 QY 1026 aaccgaccatcgaattctgcgaacacacatgcttgaagtgaagctaccaccaaaacga 1085
 Db 661 AACCCGACCATCGAATCTTCGGCAACAACAGCACTCTTTAGTGGCTACCCCAAAACGA 720
 QY 1086 tctgcttggtaccccgatgacccgtgcctttatcaccatcgtggttcccatggtgttta 1145
 Db 721 TCTGCTGTGTCACCCGATGACCGTGCCTTTATCACCATGCTGTGGTCCCATGCTGTTA 780
 QY 1146 tgaagcatatgcaatggcttcccatggtgatgaccccttgggtggtacagatgga 1205
 Db 781 TGAAGCATATGCAATGGCTTCCCATGTCATGATGCCCTTGTGTTGCTGATCAGATGGA 840
 QY 1206 caatgcaagcgcatgagactaaggagctggagtgagcctgaatgttctcggaaatgac 1265
 Db 841 CAATGCAAGCGCATGGAGACTAAGGGAGCTGGAGTGAACCTGAATGCTTGGAAATGAC 900
 QY 1266 tctggaatttgaataatgcttcaaaagcagtcattcaatgacaaaagtacaaagagaa 1325
 Db 901 TTTCTGAAGATTAGAAATGCTTAAAGCAGTCAATCAATGACAAAAGTTACAAGAGAA 960
 QY 1326 catcatgagccttccagcttcaagaagcagccgggtggagcctggaacctggccgt 1385
 Db 961 CATCATGCGCTCTCCAGCCTTCAAGAGCCGCGGTGGAGCCCTGGACCTGGCCGT 1020
 QY 1386 gttctgggtggagtttggatgaggaagcagcagcagccacacctgcccgcgagccca 1445
 Db 1021 GTTCTGGTGGAGTTTGTATGAGGCACAAAGGCGCGCCACACCTGCGCCCGCAGCCCA 1080
 QY 1446 cgaactcctcctggacagcttcccttggagcagctgattgcttcttggccgtcgt 1505
 Db 1081 CGACCTACCTGGTACCACTACCATCTTGGAGTGTGTTGCTTGGCCGCTGCT 1140
 QY 1506 gctgacagtggccttcaatcctttaaattgttcttggcttaccggaaatgcttggg 1565
 Db 1141 GCTGACAGTGGCTTCACTACCTTTAAATGTTGTGCTTATGCTACCGGAATGCTTGGG 1200
 QY 1566 gaataaaggcgagtttaagaagcccaaatccaagacccattgagaagtggtggaa 1625
 Db 1201 GAAAAAGGCGAGTTAAGAAAGCCCAAAATCCAGACCCATTCAGAAAGTGGTGGGA 1260
 QY 1626 ataagtaaaatttgacacattccctgctgattcccaacttgaacacagaatcagtg 1685
 Db 1261 ATAGGTAAAAATTTGAACCATTCCTTAGTCAITTTCAACCTGGAACAGAAATCAGTGT 1320
 QY 1686 taattcatttattcttataaggaatacttctgataaaatgaatcagccagcagatgc 1745
 Db 1321 TAAATTCATTTATTTCTTATTAAGAAATACCTTGTCAATAATTAATCAGCCCGAGAGTGC 1380
 QY 1746 tttataaaattcttataataataataatagactcgtcgtcagtaagatatttgaata 1805
 Db 1381 TTTAAAAAATTTCTTTAAATAAATAATAGACTCGCTAGTCAGTAAAGATATTGGAATA 1440
 QY 1806 tgtatgctgccccctcgggtgcttcttgatcaggtgacatgagcatttttcagagcag 1865
 Db 1441 TGTATCGTCCCCCTCTGGTGTCTTTGATCAGGATGACATGTCCTATTTTCAGAGAGC 1500
 QY 1866 tgcagaagcctggcattcagattactttcttacttgaacatgcccctgtttggag 1925
 Db 1501 TGCACAGGCTGGCATTTCTAGATTACTTTTCTTACTTGAACATGCGCTGTTGGAG 1560
 QY 1926 tgcgggattcaaggggtgccccacgcgtgcccctactgcaaatggcagtttacttat 1985

Db 1561 TCGGGATTCAAAAGGTGGTCCACGCTGCCCTACTGCAATGGCAGTTTAATCTTAT 1620
 QY 1986 ctttggctctcagatggttgcaattgataccttaaccaataatggctcagtcctcatct 2045
 Db 1621 CTTTGGCTTCTGCAGATGGTTGCAATTCATCTTAACCAATAATGGTCAGTCTCATCT 1680
 QY 2046 ctgctcctctcattagtgccacctggtgtgttttaaagaaggaagcctttgactcttag 2105
 Db 1681 CTGTCGTGTTTCATAGGTGCCACTGTGTGTTTAAAGAAGGAGGCTTTTACCTTTAG 1740
 QY 2106 agtgaagtgaaaatgaatgaatggtcgtgagtgcaacagacagatattcttctgc 2165
 Db 1741 AGTGTAGGTGAAAATGAATGAATGGCTTGGAGTGACATGAGACAGCATATGATTTCTTC 1800
 QY 2166 ttgggggaaaaagaatgatgctctgataaattggtgggtggtgtatttgagaagataaact 2225
 Db 1801 TTTGGGGA-AAAAGAAATGATCTATGAAATTTGGTGGTGGTATTGTGAGAGATAATCAT 1860
 QY 2226 tgcctatg:caaatggaagcgaattgataaaaaccccaaacacagcctatgaagtcgtg 2285
 Db 1861 TCGTTATGTCAAAATGGAGCTGAAATTTGATAAAACCCCAAAATACAGCTATGAAGTCTGG 1920
 QY 2286 gaaattactttttcttgatgatttctcacaact 2320
 Db 1921 GCAAGTTTACTTTTTTCTGATGTTTCTCACAAC 1955

RESULT 10

HUMHUGBR2

LOCUS

DEFINITION

HUMAN bilirubin UDP-glucuronosyltransferase isozyne 2 mRNA,

complete cds.

M57951.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

1. .2368

/organism="Homo sapiens"

/isolate="AK"

/db_xref="taxon:9606"

/map="Unassigned"

/clone="Z17"

/sex="female"

/tissue_type="liver"

/dev_stage="adult"

/tissue_lib="lambda-ZAP"

1. .2368

/gene="UGT2B"

30. .1634

/gene="UGT2B"

/standard_name="bilirubin UDP-glucuronosyltransferase

isozyne 2"

/BC_number="2.4.1.17"

/codon_start=1

/product="UDP-glucuronosyltransferase 2 family polypeptide

B"

/protein_id="AAA63196.1"

/db_xref="GI:184475"

/db_xref="GDB:G00-127-753"

/translation="MARGLOVPLPLATGLILLISVQPAESCKVLVYPTDGSPLSM

REALRELHARGHOAVLITFEVNHKKEKFFTLTAAYFWOKEDRVTGTCGFFE

THELLKRTSRSMALMNNVSLAHLRCCVELLHNEALIHNLNATYFDVLTDPVNCJGAV

QY 1502 tcgtgctgacagtgcccttcacatcttaataatgttgctgtatgctacaggaatgct 1561
 Db 1543 TCGTGCTGACAGTGCCTTCATCGCCTTTAAATGTTGCTATGCTACCGGAATGCT 1602
 QY 1562 tggggaataaaggcgagcttaagaagcccaataccaacccatgagaagtgggt 1621
 Db 1603 TCGGGAATAAAGGGCGAGTTAAGAAGGCCCAATAATCAGAGCCCAATTGAGAAGTTGGT 1662
 QY 1622 ggaataaagtgataattttgaacattcccttagtcatttccaaacttgaaacagagataca 1681
 Db 1663 GGAATAAAGTAAATTTGAACCATTCCTAGTCAGTCCAAATGGAACACAGATCA 1722
 QY 1682 gtgttaaatcatttattctattattagaagaatacttgcataataataacagccacaga 1741
 Db 1723 GTGTAAATTCATTTTATTC---TTAAGGAATACTTTGCATAAATAATACGCCCCAGA 1779
 QY 1742 gtgttttaaaattctcttaataataataataagactcgtagtcagtaagatatattg 1801
 Db 1780 GTGCTTTAAAAAATATCTTAAATAAATAATAAGATCACTAGTCAGTAAAGATATTG 1839
 QY 1802 aatatgtatcgccctccgctgcttgcagatgacatgctgcccatttttcagag 1861
 Db 1840 AATATGATCATG-CCCTCCGGTGCTTCGATCAAGATGACATGCCCTTCCTCAGAG 1898
 QY 1862 gacgtgcagacaggtggcatttagattactttcttactctgaacatggcctgtttg 1921
 Db 1899 GATGTCAGACAGGCTGGCGTTCTACATGCTCTTTTACTTCGGAACATGACCTGTGTTG 1958
 QY 1922 ggaatgcgggattcaaaagtgtgtccaccgctgcccactgactgcaaatggcagtttaac 1981
 Db 1959 GGAGTGCAGGATCAAAAGTGTGTCACCGCTGTCCCTATTGCATATGCGATTTAATC 2018
 QY 1982 tatctttgtgcttcagatggttgcaattgacttcaacaaataatggtcagctcctc 2041
 Db 2019 TTAATCTTTGGCTCTGTAGATGGTTGGCATTCATCCTTAACCAATAATGGTGTGCTC 2078
 QY 2042 atctctgccc-tgctctagtgccaccctgtgtgtttaaagaggagctttgttaac 2100
 Db 2079 ATCTCTGCGGTTCATAGCGGCCCTTGTGTGTTTAAAGAAGGGAACCTTGTGACC 2138
 QY 2101 tttagagctgagtgaaatgaatgaatgcttgagtgactgagacagagatatgattt 2160
 Db 2139 TTTAGACATAGTGAATGAGTGAATGCTGGAGTGCATGAGAACAGCATAGCATCT 2198
 QY 2161 ctgtctttggggaataaagatgctgctgataatgggtggtgatttattgagaagata 2220
 Db 2199 ATTGCTTTGAGGAATAAGAAATCATGCTGTAATAATTGATGGGTGTGTTTGAGAAGATA 2258
 QY 2221 atcattcttatgcataatggagctgaatttgataaaaccccaatacagctatgaagt 2280
 Db 2259 ATCATTCCTTTATGTCATAATGGAGCTGAATTTTCATAAAACCCAAATAACAGTGAAGT 2318
 QY 2281 gctgggcaagttaactttttctgtatgattttctcacaact 2320
 Db 2319 GCTGGCAACGGTA-TTTTTCCTTAATGTTTCTCTCAACT 2357

RESULT 9
 AK025403
 LOCUS
 DEFINITION Homo sapiens cDNA: FLJ21750 fis, clone COLF5999, highly similar to AF056188 Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA.
 ACCESSION AK025403
 VERSION AK025403.1 GI:10437911
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens colon mucosa cDNA to mRNA, clone_1lb:Colf
 clone:COLF5999.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)

AUTHORS Watanabe,K., Kumaqai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
 NEDO human cDNA sequencing project
 JOURNAL Unpublished (2000)
 AUTHORS 2 (bases 1 to 2003)
 TITLE Direct Submission
 JOURNAL
 COMMENT Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
 Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
 FEATURES
 source Location/Qualifiers
 1..2003
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="COLF5999"
 /clone_lib="Colf"
 /tissue_type="colon mucosa"
 /note="cloning vector pME18SFL3"
 1..2003
 /note="highly similar to AF056188 Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA"
 BASE COUNT 551 a 414 c 445 g 593 t
 ORIGIN
 Query Match 81.7%; Score 1895.8; DB 9; Length 2003;
 Best Local Similarity 98.1%; Pred. No. 0;
 Matches 1918; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
 QY 366 ttttgactatttttccaaattgcaggagttgtttaagacacacaaaattagtagaata 425
 Db 1 TCTTGACTATTATTTTCGCATGTCAGGAGTTGTTTAATGACCGAAATTAAGTAGATA 60
 QY 426 cttaaaaggagagttcttttgatgcagtggtttctcagatcccttttgataactgtggttaat 485
 Db 61 CTTAAAGGAGAGTCTTTTGTATGTCAGTGTCTTCTGGATCCTTTTTCATACCTGTGGCTTAAT 120
 QY 486 tgttgcaaatatttctccctccctccctgctgtcttcgccaggaggaactcttgcacata 545
 Db 121 TGTTCCTAAATATTTTCCTCCCTCCCTCTGTGTCCTTCACACAGGGGAATATTTTGGCACC 180
 QY 546 tcttgaagaagtgacagtgccctgctcctctctctctctctctctctctctctctctctag 605
 Db 181 TCTTGAAGAAGGTGCACAGTGCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAGG 240
 QY 606 gtctcagatgcagtgactttcaagagagagatcagacacacatcatcgtcacttgagaga 665
 Db 241 GTTCTCAGATGTCATGCTTTCAAGAGAGAGATATGGAAACCAACATCGTCGATCTGGAGGA 300
 QY 666 acattattatgcccaccgttttttccaaatgcccctagaaatagccctctgaaattctcca 725
 Db 301 CCAATTTATTTGGCAGTATCTTTTATAGAAATGCCCTAGAAATAGCCTCTGAAATTTCTCCA 360
 QY 726 aacacctgttaaggagatgactctacagccacacacatcatcttctgttctgcaacaga 785
 Db 361 AACCCCTGTCCAGGCATATGATCTCTACAGTCACACATCAATTTGGTTGTTGGCAACGGA 420
 QY 786 cttgttttggactaccacacccgtagtgcacacatgattcttctcattgttggtgataca 845
 Db 421 CTTTGTGTTTGGACTATCCCAAAACCGGTGATGCCCAACATGATCTTCAATTTGGTGTATCAA 480

```

VERSION      AF104336.1  GI:6537137
KEYWORDS     crab-eating macaque.
SOURCE       Macaca fascicularis
ORGANISM     Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
              Cercopithecinae; Macaca.
REFERENCE    1 (bases 1 to 2396)
AUTHORS      Albert, C., Vallee, M., Beaudry, G., Belanger, A. and Hum, D.W.
TITLE        The monkey and human uridine diphosphate-glucuronosyltransferase
              UGT1A9, expressed in steroid target tissues, are
              estrogen-conjugating enzymes
JOURNAL      Endocrinology 140 (7), 3292-3302 (1999)
MEDLINE      99312152
PUBMED      10385426
REFERENCE    2 (bases 1 to 2396)
AUTHORS      Albert, C., Vallee, M., Beaudry, G., Belanger, A. and Hum, D.W.
TITLE        Direct Submission
JOURNAL      Submitted (05-NOV-1998) Molecular Endocrinology, Laval University,
              2705 Laurier Boulevard, Sainte-Foy, Que G1V4G2, Canada
              Location/Qualifiers
              1..2396
              /organism="Macaca fascicularis"
              /db_xref="taxon:9541"
              60..1652
              /codon_start=1
              /product="UDP-glucuronosyltransferase UGT1A9"
              /protein_id="AAFI5346.1"
              /db_xref="GI:6537138"
              /translation="MTRTGSSPLPLCVLLLAGFAEAGKLVVVPDGHLSMRSV
              VEKLLRGHEVVVMPKVSQWOLGRPLNCTVKYTSYTLIEDLREFNAFROAOWKAQV
              QSIFLSIMSYNDIFDLFFSNCRSLFKDKLVFLKESSEDAVDFDPDTCGLIVAKY
              LSLPSVFPARGVFCVLEEGACOPAPLYPRILLGFSDAWTFKERVNHI MLEERL
              LCHREFKSALGISEILOTPVTVDLHROLSIMLLRDFVLVFPKVPWNVTFIGGIN
              CHQRPLEPFEAVINASGERGIVVFLSGSWAIEPEKKAMADALGKIPQIVLMRY
              TCTPDSLNANTILVKNLQNDLGHMPTRAFITHAGSHGIYEGICNGVPMVMPLEFG
              DQMDNKRMTKGAGVTLVNLENTSEDLNALKAVINDKRSIKENIMHLSLSLRDREVE
              PDYAVFWFEVFMHKGAPHLRPAHDLTWYOVYHSLDVGFLLAIVLTVAFLAKCCA
              YGYRCKCFKKGRVKAKHKSHT"
BASE COUNT   645 a 513 c 558 g 680 t
ORIGIN

Query Match      89.5%; Score 2076; DB 9; Length 2396;
Best Local Similarity 94.6%; Pred. No. 0;
Matches 2194; Conservative 0; Mismatches 120; Indels 6; Gaps 4;

QY  2  teagtcgagttctctgagttggtgacaggtgagcagcccttcctctatgtgtg 61
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  62  gtctgctgacgtggtgcttgcgagcaggaagactagtagtgcctcattgagtg 121
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  103  gtctgctgctggtgctggttgcgcagcaggaagctgctggtgagtcggtgag 162
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  122  ggagcactggttcacattgaggtcggtggtgagaaactcattctcaggggcatgag 181
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  163  ggagccactggctgctcactgaggtcggtggtgagaaactcattctcaggggcatgag 241
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  182  tgggtgtagtcaagcagaggtgaggtggcgaactcattctcaggggcatgag 282
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  242  agacttattcaactcattacacccctgaggtgaggtgaggttgccttgccttgcct 301
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  283  agacttattcaactcattacacccctgaggtgaggtgaggttgccttgccttgcct 342
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  302  atgctcagtgaaagcaagcagtagtatattctctatttaaggttccatcagtag 361
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  343  aggcctcaatggaaagcacaagtagtatatttctctantantagtttccctcaatg 402
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  362  acattttgactattttttcacaattgaggttgttttaagacacaaatttagtag 421
DB  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Db 230 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAATCACTGAATTCACAGTGAAGA 289
Qy 245 cttattcaactatatacaccggygagatctggaccggygagttcaaggtcttggcccatg 304
Db 290 CTCTACTCAACCTCACTACCTCTGGAGGATCTGGACGGGAATTCATGATTTGCCGGATG 349
Qy 305 ctcaatgaaagacacagtaacagatatatactctctattataatgggtttcatatacaatgaca 364
Db 350 CTCAATGAAAGCAGACAGTACAGAGTTGTTTCTCTATTTCTGAGTTCATCCAAATGGTT 409
Qy 365 tttttgaactatttttttcaactgcagagagttgtttttaaagacaaaaaatagtagaat 424
Db 410 TTTTAACTTATTTTTCGCATTCGACAGGATTTGTTAATGACCGAAATATAGTAGAAT 469
Qy 425 acttaaaagagagttctttttgatgcagtggtttctcgatccttttgataaactgtggcttaa 484
Db 470 ACTTAAAGCAGAGTCTTTTGTATGGGTGTTCTTGATCTCTTGATCCCTTGCTGGCTTAA 529
Qy 485 tigtgtccaaatattctccctccctccctccctccctccctccctccctccctccctccctccct 544
Db 530 TTGTTGCCAAATATTTCTCCCTCTCCCTCTGTTGGTCTTCCGACGGGAATAGCTTGGCACT 589
Qy 545 attctgaagaaggtgcacagtgccctgctcctctctctctctctctctctctctctctctctctag 604
Db 590 ATCTGGAAGAGGTGACAGTGCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAG 649
Qy 605 ggtctccagatgcacagctttcaagagagagatcaggaacacacacacacacacacacacacacac 664
Db 650 GTTCTCAGATGCCATGACCTTCAAGGAGAGTACGGAACACATCATGCACTTGGAGG 709
Qy 665 aaactttattatgccacccttttttcaaaaaatgcccagaaataagccctctgaaattctcc 724
Db 710 AACAATTTATTTTGGCAGTATTTTCCAAAAATGCCCTAGAAATAGCCCTGAAATTCCTCC 769
Qy 725 aaacacctgttgcagagatgaatctctcaagccacacacacacacacacacacacacacacacac 784
Db 770 AAACACCTGTCAAGCATATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAAGAC 829
Qy 785 acttggttttggactcctcaaacccggtgatgcccacacacacacacacacacacacacacacacac 844
Db 830 ACTTGTGTTTGGACTATCCCAACCCGTTGATGCCAAATATGATCTTCAATTTGGTGTATCA 889
Qy 845 actgcatacaggaagacgcttgctatgaatttgaagcctacataatgcttcttgag 904
Db 890 ACTGTCATCAGGAAGCCATTCGCTATGGAATTTGAGCCCTACATTAATGCTCTCTGAG 949
Qy 905 acatggaattgtgtttctctcttcttgggatacaatggtctcagaaattccagagaagaag 964
Db 950 AACAATGGAATGTGGTTTCTCTTTGGGATCAATGTCTCTCAAGAAATTCAGAGAAGAAG 1009
Qy 965 ctatggcaattgtgtatgctgttgggcaaaatccctccagacagctcctgtgagacacacacacacac 1024
Db 1010 CTATGGCAATTTGCTGATGCTTTGGCAAAATCCCTCAGACAGTCTCTGTGGGGTACACTG 1069
Qy 1025 gaaccgacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1084
Db 1070 GAACCCGACCATCGAATCTTGGCAACACACGATACTTGTAAAGTGGCTTACCCCAAAACG 1129
Qy 1085 acttgcttgggtccacccgtagcccgctctttatcaaccatgctggttccatgggtgttt 1144
Db 1130 ATCTGCTTGTGTCACCGATGACCGCTGCTCTTTATCACCCATGCTGTTCCTCATGCTGTTT 1189
Qy 1145 atgaacacatacgaatggcgttcccatgggtgatgatgacacacacacacacacacacacacacacac 1204
Db 1190 ATGAAGCATATGCAATGGCGTTCCTCATGGTGATGATGATGATGATGATGATGATGATGATGATG 1249
Qy 1205 acaatgcaaacgcatgagagacgaagagctggagtaacccctgaatgttctgggaatga 1264
Db 1250 ACATGCAAGCCGATGAGACTTAAAGGAGCTGGAGTACCCCTGAATGTCTTGGAAATGA 1309
Qy 1265 ctcttgagatttagaaatgctctaaagcagctcatcaatgacaaaagttaacagggaga 1324
Db 1310 CTCTGGAAGATTTAGAAAATGCTCTAAAGACAGTCAATCAATGACAAAAGTTACAGGAGA 1369

Qy 1325 acatcatgctcctccagccttcacaagggaccgcccgggtggagccgctggacctggccg 1384
Db 1370 ACATCATCGCTCTCTCCAGCTTCAAGAGGACCGCCGGTGGAGCCGCTGGACCTGGCCG 1429
Qy 1385 tgttctggctggagtttgatgaaggacaaagggcgcgcaacacctgccccgcgcccc 1444
Db 1430 TGTCTGGCTGGAGTTGTGATGAGGACAAAGGGCGCGCACACCTGCGCCCGCCGACGCC 1489
Qy 1445 agcaactcectggctacacagttaccattccttggacgtgattggtttctctcttggccgtcg 1504
Db 1490 ACCACTCTACCTGTACACAGTACCATTCCTTTGGACGTGATTGTTCTCTTGGCCGTG 1549
Qy 1505 tctgcaactggccttcacacactttaaatgttgtcttatggctacccggaaatgttgg 1564
Db 1550 TGTCTGACAGTGGCTTCTACACCTTTAAATGTGTGTCTTATGGCTACCGGAAATGTCTGG 1609
Qy 1565 ggaataaagggcgagtttaagaaagcccaaaatccaaagacccttgagaagtggtggga 1624
Db 1610 GGAAAAAAGGGCGAGTTAAGAAAGCCACAAATCCAAAGACCCATTGAGAAAGTGGTGGA 1669
Qy 1625 aataaggttaaaattttgaaccattccctctgattcatttccaaacttgaacacagaatcagtg 1684
Db 1670 AATAAGGTAAATTTTGAACCATTCCTCTAGTCAATTTCCAAACTTGAHAACAGAAATCAGTG 1729
Qy 1685 ttaattcatttattcttatttaaggaataactttgcataaaataatcagccccagtg 1744
Db 1730 TTTAAATTCATTTTATTTTAAAGGAAATACTTTGCATAAAATTAATCAGCCCGCAGAGTG 1789
Qy 1745 ctttaaaaaatctcttaataaaaaataatagactcgtagctagctgaagataatttgaat 1804
Db 1790 CTTTAAAAATTTCTTAAATAAAAAATTAATAGACTCGTAGTCAATAAGATATTTGAAT 1849
Qy 1805 atgtatcgtcccccctccggtgtctttgatcagagtgacatgtgccatttttcagaggac 1864
Db 1850 ATGTATCTGTCGCCCTCCCGGTGTCTTTGATCAGAGTACATGTGCCATTTTTCAGAGGAC 1909
Qy 1865 gtgcagacagctggcgatctctagattcttcttactctgaacatggcctgttggga 1924
Db 1910 GTGCAGACAGCTGGCATCTTAGATTACTTTCTTACTCTGAACATGGCTCTGTTTGGGA 1969
Qy 1925 gtgcggagttcaaaaggtgccccccgctgccccctactgcaaaatggcagttttaaacttta 1984
Db 1970 GTGCGGGAATTCAAAAGGTGTGCCACCGCTGCCCTACTGCAAAATGGCAGTTTAAATCTTA 2029
Qy 1985 tcttttggctcttcgcagatgggttgcgaattgatc- cttaaaccaataatggctacat 2043
Db 2030 TCTTTTGGCTTCTGCAGATGGTTGCAATTTGCAATTTGCAATTAATGGTCACTCCCTCAT 2089
Qy 2044 cctctgctcttcacagtgccacctgtgtttaaagaagggaaggtttgtaccttt 2103
Db 2090 CTTCTGCTCTCTATAGTGGCCCTTTGTGTGTTTAAAGAGGGAAGCTTTGTACTTTT 2149
Qy 2104 agagtgtagtgaatgaatgaatggcttggagtgacactgagacacacacacacacacacacacac 2163
Db 2150 AGAGTGTAGTGAATGAATGAATGAATGGCTTGGAGTGCACGTGAGAACAGCATATGATTTTTT 2209
Qy 2164 gctttgggaaaaagaatgactatgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2223
Db 2210 GCTTTGGGAAAAAAGAAATGCTATGAAATTTGGTGGTGGTGTATTTGAGAAAGATAATC 2269
Qy 2224 attgcttatgtcaaatggagctgaatttgataaaaaaccccccaaaatcacagctatgaagtgt 2283
Db 2270 ATTGCTTATCTCAATGGAGCTGAATTTGATTAATAAACCCAAATACAGCTATGAAGTGTCT 2329
Qy 2284 gggcaagttacttttttttctgatgttttctcaact 2320
Db 2330 GGGCAAGTTTACTTTTTTTTTTGTGTTTCTCTCCAAT 2366

RESULT 6
BC020971
LOCUS

2423 bp mRNA linear PRI 22-JAN-2002

Qy 1205 acaatgcacagcgtgagactaaaggagctgagctgagctgacccctgaattgtctggaatga 1264
Db 1220 ACAATGCAAGCGATGAGACTAAGGAGGCTGAGTGCACCTGAATGTTCTGGAATGA 1279
Qy 1265 ctctgaagatttagaataatgctctaaagcagtcacatcaatgacaaaagtacaaagaga 1324
Db 1280 CTCTGAGAGATTAGAAATGCTCTAAAGCAGTCAATCAATGACAAAGATTACAAGAGA 1339
Qy 1325 acatcatgagctctccagcctccaaagacaccccggtgagcgctggaacctggccg 1384
Db 1340 ACATCATGCGCTCTCCAGCCTTCACAAGGACCGCCCGTGGAGCGCTGGAGCTGGCGG 1399
Qy 1385 tgtctggtggtgagttgtgtaggcacaaaggcgccgcacacacctgcccccgagccc 1444
Db 1400 TGTCTGGGTGGAGTTTGTGATGAGGCACAAAGGCGGCCACACCTGCGCCCGAGCCG 1459
Qy 1445 acaacctcaactggtaccagtgaccattcccttgagctgagctgagtttctctctggccgtg 1504
Db 1460 ACAGCCTCACCTGTTACAGTACCATCTCTTGGACGTGATGTTGTTCTCTTGGCGGTGG 1519
Qy 1505 tgcagcagtgcccttcacacctttaaatgtgtcttatgctaccggaatgcttg 1564
Db 1520 TGTGACAGTGGCTTCATCACCCTTAATGTGTCTTATGCTTACCGGAAATGCTTGG 1579
Qy 1565 ggaataagggcggttaagaagcccaaaatccaaagacccatctgagaagtgggtggga 1624
Db 1580 GGAATAAAGGGCGAGTTTAAAGAACCCCAAAATCCAAAGACCCATTTGAGAAGTGGGTGG 1639
Qy 1625 aaataaggtaaaatttgaacacattccctagctagctattcccaacctgaaaaacagatcagtg 1684
Db 1640 AATAAGGTAATAATTTGAACCATTCCTTAGTCATCTTCCAACTCTCAAAACAGAAATCAGTG 1699
Qy 1685 taaattcatcttattctatttaagaagaatacttgcataaaatgaatcagccccagtg 1744
Db 1700 TTAATTCATTTATTTCTTAAAGGAAATATCTTGCATAAATTAATCAGCCCCAGAGTG 1759
Qy 1745 ctttaaaaaattctcttaaaaaataatagacgctgctagctagctaaagataatttgaat 1804
Db 1760 CTTTAAAAATTTCTTAAAAATAAATAATAGACTGCTAGTACGTAAGATATTGTAAT 1819
Qy 1805 atgatatgtgccccctccggtgtctttgatcagagatgaatgacatgtgcaatttttcagaggac 1864
Db 1820 ATGATATGCTGCCCCCTCCGCTGCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1879
Qy 1865 gtgcagacaggtggtgcttctagattacttttctgctgaaacatgctgctgttggga 1924
Db 1880 GTGCAGACAGGCTGCGCATTTACTTTCTTACTCTGAAACATGCGCTGTTTGGGA 1939
Qy 1925 gtgcgggattcaaaagggtgtccacacgctgccccctactgcaaatggcagttttaattctta 1984
Db 1940 GTGCGGATTCAAAAGGTGTGCCACCGCTGCCCTACTGCAATGGCAGTTTAAATCTTTA 1999
Qy 1985 tctttgctctcagatggttgcaattgacatccttaacaaatgaatgctcagctccatc 2044
Db 2000 TCTTTTGGCTTCAGATGTTGCAATGATGATCCTTAACCATATATGTCAGTCCCTCATC 2059
Qy 2045 tctgtctgcttcattaggtgcccacacctgtgtgtttaaagaaggagctgttgaccttta 2104
Db 2060 TCTGTCTGCTTCATAGTGTGCCACCTTGTGTGTGTTAAAGAGGAAAGCTTGTGACCTTTA 2119
Qy 2105 gagtgtaggtgaatgaatgaatggtgtgagtgacatgagaaacagcatatgattcttg 2164
Db 2120 GAGTGTAGGTGAATGAATGAATGCTTGGAGTGCATGAGAAACAGCATATGATTTTTTG 2179
Qy 2165 ctttgaggaaagaatgatgctatgaataattggtggtggtgtgtatttgagaagaataatca 2224
Db 2180 CTTTGGGAAAAAGAAATGATGCTATGAAATTTGGTGGGTGGTGTATTTTGAAGAATATATCA 2239
Qy 2225 ttgcttatgtcaaatggagctgaatttggataaaaaacccaaataacagctatgaagtgtcg 2284
Db 2240 TTGCTTATGTCAATGGAGCTGAATTTTGATAAAACCCAAATACAGCTATGAAGTGTG 2299

Qy 2285 ggaagttacttttttttttttttttttttttttttttttttttttttttttttttttttt 2320
Db 2300 GGCAAGTTTACTTTTTTTTTTTTTTTTTTTTTTTTGTATGTTTCTCCCAACT 2335
RESULT 5
LOCUS AF462268 2394 bp mRNA linear PRI 07-FEB-2002
DEFINITION Homo sapiens UDP-glucuronosyltransferase UGT1A8*1 (UGT1A8) mRNA,
UGT1A8*1 allele, complete cds.
ACCESSION AF462268
VERSION AF462268.1 GI:18568218
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Huang,Y.-H., Nguyen,N., Geske,D. and Tukey,R.
TITLE Identification and functional characterization of UGT1A8*1,
UGT1A8*2 and UGT1A8*3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2394)
AUTHORS Huang,Y.-H., Nguyen,N., Geske,D. and Tukey,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Pharmacology, UCSD, 9500 Gilman Drive, La
Jolla, CA 92093, USA
FEATURES
Location/Qualifiers
1..2394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="Colon"
1..2394
/gene="UGT1A8"
/codon_start=1
/product="UDP-glucuronosyltransferase UGT1A8*1"
/protein_id="AA175964.1"
/db_xref="GI:18568219"
/translation="MARTGWTSPICVLSLLTTCGPAENAGKLLVYVMDGSHWFTMQSV
VERLLLRHEVYVVMVEVSNQSLNCTVKTYSTSTYLEDLDFDFDAQMKQAV
RSLSLFTSSNGFNFLDFSHCRKLVKLFKESSDFAVFLDFDAGCLIVAKY
FSLPSVFEARGIAHYLEGAOCAPLSPVPRILLGLFSADMTFKERVRNHIHLEBHL
EQOYSKNALEIASEIILOTPTVAYDLYSHTSINLLRDTFVLDYPRKVPNNIMFTGGIN
CHQGPLMFEFEAVINASGEHIVFSGSMVSEIPEKKAMAIADALCKIPOTVIMRY
TGTSPNLNNTILYKWLQNDLLGHPMTAFITHAGSHGVYESICNGVPVWMPLEF
DOMNAKMETKGAGVTLNVLKMTSELENALKAVIDNSKSYENLRLSSLHKDRPE
PDLAVFWVEFVHRHKGAPHLRPAADHLITWYQVHSDVDVIGFLAVLLVAFITPKCA
YGYRCLGKGRVKKHKKSTH"
BASE COUNT 624 a 513 c 555 g 702 t
ORIGIN

Query Match 95.1%; Score 2207.4; DB 9; Length 2394;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2255; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
Qy 5 gctgcagttctctgagcttgacagaggtggaccagccccctctctctatgtgtgtc 64
Db 50 GCTGCAGTTCTCTCATGCTGTCGACAGGTGGACAGCCCATTCCTCTATGTGTTCTC 109
Qy 65 tgcctgacctgtgcttgccgagcagcaggaactgagtagtgcctatgagtaggga 124
Db 110 TGCTCTGACCTGTGGCTTGTCTGAGGAGGAGGCTGCTGAGTGCCTCATGTTGGGA 169
Qy 125 gccactggttcaacatgagctggtggtgagaaactattctcagggggcagtaggtg 184
Db 170 GTCAGTGTTCACCATGCAGTCGGTGGTGGAGAACTTATCTCAGGGGCGATGAGTGG 229
Qy 185 ttgtagtcatgcagaggtgagttggaactgggaagatggaactggaattgacagtaga 244
|||||

Db	1021	ACTGGAAACCGGACCATGAATCTTGGAAACAACAGATACTTGTAAAGTGGCTACCCCAA	1081
Qy	1081	aacgatctgttggtaccacccgatgacccgtgcctttatcacccatgtggttcccacatggt	1140
Db	1081	AACCATCTGCTTGTGTACCCGATGACCGGTGCTTTATCACCCATGCTGGTTCCTCATGTT	1140
Qy	1141	gtttatgaagcatatgcaatgcgttcccctagtatgacccctgttttgatcacg	1200
Db	1141	GTTTATGAAGCATGCAATGGCTTCCCATGCTGATGATGCCCTTGTGTGGATCAG	1200
Qy	1201	atggacaatgcaaaagcgcctgagagactgaaggagctggagtgacccctgaaatgtcttggaa	1260
Db	1201	ATGGACAATGCAAAAGCCGATGGAGACTAAGGAGCTGGAGTGACCTGGAATGCTTGGA	1260
Qy	1261	atgactcttgaagattttagaaaatgctctaaaagcagtcataatgacaaaagtttacaag	1320
Db	1261	ATGACTTCTGAAGATTTAGAAAATGCTCTAAAGCAGTCTCATATGACAAAAGTTTACAG	1320
Qy	1321	gagacacatcgcctctccagctccacaaggaacgcgcggtgagcgcgtggaacctg	1380
Db	1321	GAGACATCTGCGGCTCTCCAGCCTTCAACAAGACCGCCGCTGGAGCGCTGGACCTG	1380
Qy	1381	gcctgtcttgggtggagtttgtgatggagcacaaggcgcgccacacctgcgcgccgcga	1440
Db	1381	GCCGTCTTCTGGGTGGAGTTTGTATGAGCACAAGGGCGCCACACCTGCGCCCGCCGA	1440
Qy	1441	gccacgacctcaccgtgtaccagtaaccattccttgacgctgattggttctctctggcc	1500
Db	1441	CCCCACGCTCCTCTGTATACCATGACCATTCCTTGGACGTGATTGGTTTCTCTTGGCC	1500
Qy	1501	gtcgtctgacagtggcctctcatcactttaaatgttgtgtctatggctacacggaatgc	1560
Db	1501	GTCGTCTCAGACGTGGCTTCATCACTTTAAATTTGTGTCTTATGCTTACCGGAATGC	1560
Qy	1561	tgggaaaaaaggcgagtaagaagaacccacaataccaagaccatgagaagctgggt	1620
Db	1561	TTGGGAAAAAAGGCGAGTTAAGAAAGCCCAATCCAAGACCATTTAGAAGTGGGT	1620
Qy	1621	ggaaaaaagtgaaaaattttgaacatttccctagtcatttccaaacttgaaaacagaaac	1680
Db	1621	GGAAATAGGTAAAAATTTTGAACCATTTCCCTAGTCAATTTCCAACTTTGAAAACAGAA	1680
Qy	1681	agtgtaaatcattcttattcttataaggaatactttgtcataaatatcaatcagccccag	1740
Db	1681	AGTGTTAAATTCATTTTATTTCTTATTAAAGGAATACTTTGCTAATAATTAATCAGCC	1740
Qy	1741	agtgctttaaaaaattctcttaataaaaaataaagactcgtagtcagtaagaatattt	1800
Db	1741	AGTGCTTTAAAAAATTCCTTTAAATAAAAAATAAATAGATTCGCTAGTCAGTAAGATAT	1800
Qy	1801	gaatagtatcgtgcccccctcggtgtcttgtatcagtagtacatgtgccatttttcaga	1860
Db	1801	GAATATGATCTGTGCCCTCGCGGTCTTTGATCAGATGACATGTGCCATTTTTCAGA	1860
Qy	1861	gaacgtgcagacagctggcatcttagattactttcttactctgaaacatggcctgttt	1920
Db	1861	GGACGTGCAGACAGGTGGCATTTAGATTACTTTTCTTACTCTGAACATGGCTGCTTT	1920
Qy	1921	ggagtgcggaattccaaagtggttcccacacgctgcctcactgcgaatgagcagtttcaat	1980
Db	1921	GGGAJTGCGGGATTCAAAAGTGTGCCACCGCTGCCCTCTACTGCAATGGCAGTTTTTA	1980
Qy	1981	cttaacttttggtctctgcagatggttgtgcaattgatactcttaaccaataatggctagtc	2040
Db	1981	CTTATCTTTTGGCTTCTGCAGATGGTTGCAATTTGATTCCTTAAACAATAATGGTCAGT	2040
Qy	2041	catctctgctcgtcttcatagtgccaccttgtgtttaaagaagggaagcctttgtacc	2100
Db	2041	CATCTCTCTCTCTCATAGTGCACCTTGTGTGTTTAAAGAAAGGAGCTTTGTATC	2100
Qy	2101	tttaagtgtagtgaaatgaatgaatgccttgagctggagctgagaacagcatatgattt	2160
Db	2101	TTTAAAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCATCTGAGAACAGCATATGAT	2160

QY	2161	cttgcttgggaaaaagaatgcatagaattgggtgggtgttatgttgagaagata	2220
Db	2161	CTTGTCTTTGGGAAAAAGAAATCATGCATTATCAAAATTTGGTGCGTCGTATTTTGAGAAGATA	2220
QY	2221	atcattgcttatgtccaatggagctgaattgatataaaacccaaaatacacgactatgaact	2280
Db	2221	ATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAARCCAAAATACACTATGAAC	2280
QY	2281	gtggggcaagttaacttttttttcctgatgttttctctacaact	2320
Db	2281	GCTGGCAGGTTTACTTTTTCTGATGTGTTCTCTACAAC	2320

RESULT 3
AF0561188 LOCUS DEFINITION
Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA, complete cds.
AF0561188
AF0561188.1 GI:3414797
human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2320)
Ciotti,M., Potter,C. and Owens,I.S.
JOURNAL Human phenol metabolizing UDP-glucuronosyltransferase Unpublished
AUTHORS 2 (bases 1 to 2320)
Ciotti,M., Potter,C. and Owens,I.S.
JOURNAL Direct Submission
SUBMITTED 30-MAR-1998 Heritable Disorders Branch, NICHD, 9000 Rockville Pike, Bethesda, MD 20892, USA
AUTHORS 3 (bases 1 to 2320)
Ciotti,M., Potter,C. and Owens,I.S.
JOURNAL Direct Submission
SUBMITTED 12-AUG-1998 Heritable Disorders Branch, NICHD, 9000 Rockville Pike, Bethesda, MD 20892, USA
REMARK Sequence update by submitter
COMMENT On Aug 13, 1998 this sequence version replaced gi:3025895.
FEATURES Location/Qualifiers
1..2320
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q37"
 /tissue_type="liver"
 1..2320
 /gene="UGT1A9"
 19..1611
 /gene="UGN1A9"
 /codon_start=1
 /product="UDP-glucuronosyltransferase 1A9"
 /protein_id="AAC31425.1"
 /db_xref="GI:3025896"

gene /translat="NACGTMTSPLCVLLCTLCTGAEAGKLVVPMDSGHMTRSR VSKITLHGSEVNDIFDLPFNSGRSLFKDKTVSTYSVTLEDLDRFAFAHAQKAQV RSLTSVLMSGYNDIPDLFNSGRSLFKDKTVKLSVESDFDAVDFPDNCGLIVAKY FSLSVFARGILCHVLEEGACQAPLSVPVRILLGFSDAMTFERVNRNIHLEHL LCHRFKNALTEASIILOTPVTDYLISHTSWLLRTDFVDLDIPIKPVNNFIIGIN CHQKPPLMEFTAFIINASGEHGVLFPSGSWSEIPEKKAMAIDAAGKIPOTVLMRY TDOMNKRMETGAGCVTLNVNLTSSBDLENALKAVINDSKENIMRLSLSLHKDRPFG PDLIAVFPVEFWRHKGAPHLRPAADHLTWTOYHSLDVICFLLAWLTVATFITFKCA YCYRKCLGKRGVRKKAHKSXTH"

BASE COUNT 608 a 495 c 547 g 670 t
ORIGIN


```

|||||
Db 1861 GGAGCTGCAGACAGCTGGCATTCTAGATTACTTTCTTACTCTGAACATGGCTGTTT 1920
Qy 1921 gggagtgaggatcaagtggtccacccgctccctactgcaaatggagtttaaat 1980
Db 1921 GGGAGTGGGGGATTCAAGGTGGTCCACCGCTGCCCTACTGCAATGGGAGTTTAAAT 1980
Qy 1981 ctatctcttgctctgcagatgggtgcaattgatcccttaaccataaatggtcagtcct 2040
Db 1981 CTTATCTTTTGGCTTCTGCAGATGGTGTGCAATGTGATCTTTAAACCAATAATGTCAGTCT 2040
Qy 2041 catctctgctctcatagtgccaccttgatgttttaaaagagaaagctttgtacc 2100
Db 2041 CATCTCTGCTCTTCATAGGTGGCCCTTGCTGTGTTTAAAGAGGAGCTTTGTACC 2100
Qy 2101 tttagtgtagtgaatgaatgaatggcttgagtgactgagacacgcatatgattt 2160
Db 2101 TTTAGAGCTAGGTGAATGAATGAATGGCTTGGAGTGCACCTGAGAACAGCATATGATTT 2160
Qy 2161 ctgcttgaggaaagaaatgatctatgaattggtggtgtatttgagaagata 2220
Db 2161 CTGCTTTGGGAAAAGAAAGATGATGCTATGAATTTGGTGGTGGTGTATTTGAGAAGATA 2220
Qy 2221 atcattgcttatgtcaaatggagctgaatttgataaaaaaccccaatacagctatgaagt 2280
Db 2221 ATCATTTGCTTATGCTCAATGAGCTGAATTTGATAAAAAACCCAAAATACAGCTATGAAGT 2280
Qy 2281 gctggagcaagtttactttttcttgatgtttctcacaact 2320
Db 2281 CTTGGGCAAGCTTTACTTTTTTCTGTAGTGTTCCTACAACT 2320

RESULT 2
AX286764
LOCUS AX286764 2320 bp DNA linear PAT 21-NOV-2001
DEFINITION Sequence 3 from Patent WO0180896.
ACCESSION AX286764
VERSION AX286764.1 GI:17048799
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Ratalin,M.J., Innocenti,F. and Iyer,I.
TITLE Flavopiridol drug combinations and methods with reduced side
effects
JOURNAL Patent: WO 0180896-A 3 01-NOV-2001;
ARCH DEVELOPMENT CORPORATION (US)
FEATURES
source
1..2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
19..1611
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD12275.1"
/db_xref="GI:17048800"
/translation="MACTGWTSPILCVLLTCGFABAGKLLVPMDSGSHFTMRVS
VEKILRGHEVVMPEVSWQLGRSLNCTVKYVSTYLEDLDRFKAFAHQWRAQV
RSIYLLRGHEVVMPEVSWQLGRSLNCTVKYVSTYLEDLDRFKAFAHQWRAQV
FSLPSVVFARGILCHVLEBGAQCPAPLSVYPRILLGFSDFAMTFKERVNRHIMLEHL
LCHREFKNALFASLTIOTPTVEYDLYSHTSIWLRTDFLDYKPKVPMNMFIFGSLN
CHQKPLPMFEAFYINASGEHGIYVFLSGMSWSEIPEKKAMADALGKIPQVLMRY
TGTNRSLNANTLLVKNLPQNDLGHFWTRAFITIHAGSHGVSEICNGVPVWMPLEFG
DOMNARKMETRGAVTLNLEMTSEDLNAUKAVINDKSYENIKENIKLSLHKDRPE
PLDLAVFWFEVFMHKGAPHLRPAADHLTWYQYHSLDWIGFLVALVLTVAFTFKCA
YGYRKLCKGKRGVKAHKSHT"
BASE COUNT 608 a 495 c 547 g 670 t
ORIGIN
|||||

```

Query Match 100.0%; Score 2320; DB 6; Length 2320;

```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ctaactacattctctgagcttgacacaggtgagagagccctctctctatggtg 60
Db 1 CTGAGCTGAGTCTCTGATGGCTTGACAGGGTGGACAGAGCCCTTCCCTCATGTG 60
Qy 61 tgtctgtctgactctggcttgcagagggagggagctactggtgagtgccatggt 120
Db 61 TGTCTGCTGCTGACTGTGCTTGGCAGGCGAGGGAAGCTACTGCTAGTGCCTCAT 120
Qy 121 gggagccactggttcacatgaggtcgtggtggagaaactcatctcaggggcatgag 180
Db 121 GGGAGCCACTGGTTCACCATGAGGTGGTGGTGGAGAACTCATCTCAGGGGCAT 180
Qy 181 gtgggttagtcatgcccagaggtgagttggaaactgggaagatcaactgaattg 240
Db 181 GTGGTGTAGTCACTGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATGG 240
Qy 241 aagacttatcaactctatataccctggagatctggaccgggagttcaaggcttt 300
Db 241 AAGACTTATCAACTTCATATACCTGGAGGATCTGGACGGGAGTTCAAGGCTT 300
Qy 301 catgctcaatggaaagacaaagtacaaagtatatctctatataatgggttcaca 360
Db 301 CATGCTCAATGGAAAGCACAAAGTACGAAGTATATATTTCTATTAAATGGGT 360
Qy 361 gacattttgacttatttttcaaatgacagaggttgttttaagacacaaaaatt 420
Db 361 GACATTTTGTGCTTATTTTTCAAATTCAGAGGATTTGTTTAAAGACAAAAA 420
Qy 421 gaatactaaaggagaggtcttttgatgcagtggtttctcgatcccttttgata 480
Db 421 GAATACTTAAAGAGAGAGTCTTTTGTATGAGTGTTCGATCCTTTTGATAACT 480
Qy 481 ttaattggtgcaaatattttctccctccctcgtggtcttcgagggaggaatact 540
Db 481 TTAATTGTTGCCAAATATTTTCCCTCCCTCCGTTCTTCGCCAGGGAATACT 540
Qy 541 caatacttgaaagaggtgcaagtgccctgctctctctctctctctctccagaa 600
Db 541 CACTATCTTGAAGAAGGTGGACAGTGCCTGCTCTCTCTCTCTATGTCGCCA 600
Qy 601 ttgggttctcagatgcatgacttcaaggagaggtacggaaacacacatcatgca 660
Db 601 TTAGGTTCTCAGATGCCATGACTTTTCAAGAGAGAGGTACGGAACCATCATG 660
Qy 661 gaggaacatttatatgccaccttttttcaaaaatgccctagaaaatagctctga 720
Db 661 GAGAACATTATATATGCCACCGTTTTCAAAATGCCCTAGAAATAGCCTCTGA 720
Qy 721 ctcaaacacctgtgtacggagatgatctctacagccacacatacaattggttgc 780
Db 721 CTCCAACACCTGTGTACGAGTATGATCTCTACAGCCACACATCAATTTGGT 780
Qy 781 acggaactgttttgactatcccaacccgtgatgcccaacatgatctctatggt 840
Db 781 ACGACTTGTTTTGGACTATCCCAACCCGTGATGCCCAACATGATCTTCAT 840
Qy 841 atcaactgcatcagggaagccggttgctatgggaatttgaagccctacataatg 900
Db 841 ATCAACTGCCCATCAGGAAGCCGCTTGCCTATGGAATTTGAAGCCTACATTA 900
Qy 901 ggaagacatggaattggttttctcttcttggatcaatgggtctcagaaaattcc 960
Db 901 GGAGACATGGAATTTGTGTTTTCTCTTTTGGATCAATGGTCTCAGAAATCC 960
Qy 961 aaagctatgcaattgctgattgctttggggcaaaaatccctcagacagctct 1020
Db 961 AAAGCTATGCAATTTGCTGATGCTTTGGGCAAAAATCCCTCAGACAGTCT 1020
Qy 1021 actggaaccgaccatgaactcttgcaaacacagatacttgttaagtggctacc 1080
Db 1021 ACTGGAACCGACCATGAATCTTGCAAAACACAGATACCTTGTAAAGTGGCT 1080

```



```

/codon_start=1
/protein_id="GAD12274.1"
/db_xref="GI:17048738"
/translation="MAGTQVSLPCLVCLLLTCGFAEAGKLLVPMDSGHWFMTSRV
VEKLIRGEVVMVPMQWGLRSNCTVYTSYLLDLDREFKAFARQAQV
RSYLLMGSNIDFLFNSCRFSPKDKLVLEYLKESFDVFLDPDNGCLIVAY
FSYSPFARGILCHYLEEGCAPGLSPVRLIKESFDAMTFKVRVNHIMLIEHL
CHRFKFNALTEAIIQLTQTEYDLSHTSYIMLLRDVLDVYKPMVNMPIFGGIN
CHQKPLMEFAEYINAGSGHGVTVFSLGSGVSPKPKAMAIDALCKIPOTVLVRI
TGTRPSNLMNTIKVLPWLDNLLGHPTAFITHAGSGHGVSTICGLSSVPMFLG
QDMNKANMNTKAGAVTLNLYEMTSDLENLAIRINDKSYKENTIRMLSLHKDRPE
PLDVAIVYEFVPMQWGHKAPLRPAALHDLTWYQYHSLDVIGFLVAVLVIAFTFKCA
YGYBKLCCKGPPKKKSKYTH"

```

Sample	Weight (g)	Volume (ml)	Concentration (g/ml)
a	495	547	670

Query Match	100.0%	Score 2320;	DB 6;	Length 2320;
Best local similarity	100.0%;	Pred. No. 0;		
Matches 2320;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ctcagctgcagttctctctgattgctgcacaggggtgcacagccctctctctatgtgtg	60	
DB	1	CTCAGCTGCAGTTCTCTGATGCTTGCACAGGGTGGACAGCCCTTCTCTATGTGTG	60	
QY	61	tgtctgctgtaacctgacctggctttccqagagcgaggaaagctactgtagtgcacatggat	120	
DB	61	TGTCGTGCTGACCTGTGGCTTTCCGAGGACAGGAAGCTACTGGTAGTGCCCATGGAT	120	
QY	121	ggagagccactggttcaaccatgaggtcggtggtggagaaactcattctcaggggggcagtag	180	
DB	121	GGGAGGCCACTGGTTCCACCATGAGTGGGTGGTGGAGAAACTCATTTCTCAGGGGGCATGAG	180	
QY	181	gtggtgtgagtcatgcacaaagtgagttggcaactgggaagatcaactgtaattgcacagtg	240	
DB	181	GTGGTTGTGATCTACGCCAGAGGTGAGTTGGCACTGGGAAGATCACTGAAATGGCACAGT	240	
QY	241	aagactta tcaaaccttcataaccctggagagctctggaccggggagttcgaagcctttggcc	300	
DB	241	AAGACTTATTCAACTTATATACCTTGGAGGATCTGGACGGGAGTTCAAGGCTTTTGCC	300	
QY	301	catgctcaatggaaagcacaaatgacagatatactcttattaatgggtttacataaat	360	
DB	301	CATGCTCAATGGAAGACACAGTTACGAAGTATATATTCATTAATGAGGTTCACTACAAT	360	
QY	361	gacacttttgacttatttttttcaaatggcaggagttggtttaagacacaaaaa ttagta	420	
DB	361	GACATTTTGTGACTATTATTTTTCAAATGGCAGGAGTTGTGTTAAAGACAAAAAATTAGTA	420	
QY	421	gaatacttaaaagagagtgctctttgatgcagtggtcttcgcactctttgataactgtggc	480	
DB	421	GAATACTTAAGGAGAGTCTTGTATGACAGTGTGTTCTCGATCCTTTTGTAACTGTGGC	480	
QY	481	ttaattgttcgcaaatattctccctccctccgttggtcttcgcaggggaaatactttgc	540	
DB	481	TTAAATTGTGGCAAAATATTTCTCCCTCCCTCCGCTGGTCTTCGCCAGGGGAATACTTTGC	540	
QY	541	cactatcttgaagaaggtgcacagtgccctgcctctctcttctctatgtccccagaaattctc	600	
DB	541	CACATATCTTGAAGAAGGTGCACAGTGCCCTGCTCTCTTCTTCTATGTGCCCAAGAATCTC	600	
QY	601	ttagggttctcagatgccatgacttcaaggaaagagtagtcaggaacacacatcatgcaattg	660	
DB	601	TTAGGGTCTCAGATGCCATGACTTTCAAGGAGAGAGTAGCGGAACCACTATCATGCACCTG	660	
QY	661	gaggaaacatttatatgccaccgctttttcaaaaaatgcctcagaaa tagcctcttgaaatt	720	
DB	661	GAGSAACATTTATTATGCGACGGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAATTT	720	
QY	721	ctccaaaacactgttacggagtagtatctctacagccacacacatacaatttgggtgtgcga	780	
DB	721	CTCCAAACACTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGTTGTTGCGA	780	

ORIGIN

Query Match 27.8%; Score 644; DB 10; Length 723;
 Best Local Similarity 96.7%; Pred. No. 1.1e-157;
 Matches 700; Conservative 0; Mismatches 20; Indels 4; Gaps 4;

QY 1078 caaaacgctctgcttgtaacccgcatgaccgctgctcttataccacatgctgttcccat 1137
 Db 2 CAAGAGATCTGCTGTGTACCCGATACCCGCTG-CTTATACCATGCTGTGCTCCCAT 60
 QY 1138 ggtgttttgaagacatgcatgcaatggcttcccatggtgatgatgcttctgttggat 1197
 Db 61 GGTGTTTATGAACATATGCAATGGCTTCCCATGCTGATGATGATGCTGCTGCTGAT 120
 QY 1198 cagatgagacaatcaagagcagatgagactaagggagctggagtagaccctgaatgtctg 1257
 Db 121 CAGATGGAATGCAAGAGCGATGGAGACTAAGGGAGCTGGAGTGAACCTGAATGCTTG 180
 QY 1258 gaaatgactctgaagattagaataatgctctaaagcagtcataatgacaaaattac 1317
 Db 181 GAAATGACTTCTGAAGATTAGAAAAATGCTTAAAGCAGTCTCATCAATGACAAAATTAC 240
 QY 1318 aaggaacacatcagcctctccagcttcaacagagcagcccggtgagcgcgtgagac 1377
 Db 241 AAGGAGACATCATGCGCTCTCCAGCCTTCAAGAGACCCGCGGTGGAGCGCTGGAC 300
 QY 1378 ctggccgtgtctctggtgagttgtgatgagcacaagggcgccacacacctgcgcccc 1437
 Db 301 CTGGCCGTGTCTGGTGGAGTTTGTGATGAGGCACAAAGGGCGGCCACCTGGCCGCC 360
 QY 1438 gcaagccagcactcaccctggtaccagtagcattctctgagctgattgttctcttg 1497
 Db 361 GCAGCCACAGACCTACCTGGTACCAGTACCATTCCTTGGAGCTGATGTTGCTCTTG 420
 QY 1498 gcctgtgtgacagtgccctctacacctttaaattgtgtgtcttatgctaccggaaa 1557
 Db 421 GCGCTCTGCTGACAGTGCGCTTCATACCTTTAAATGTTGTGCTTATGCTACCGGAAA 480
 QY 1558 tctctgggaaaaagggcagtttaagaaagcccaaatcaagaccattgagaagt 1617
 Db 481 TCGTTGGGAAANAGGGCGAGTTAAGAAAGCCCAAAATCCAGACCCATTGAGAAGTG 540
 QY 1618 gttgggaataaagttaaaa-ttttgaaccattcccttagtcattttccaaacttgaaacag 1676
 Db 541 GGTGGGAAATAAGTAAAAATTTTGAACCATTCCTTAGTCAATTCCTCAAACTTGAAACAG 600
 QY 1677 aatcagtttaaatctatttattcttataaggaataactttgcataaataatcaagcc 1736
 Db 601 AATCAGTCTTACATTCATTTATTCTTATTAAGGAAATAC-TTGCATAAATTTATCAG-C 658
 QY 1737 caagtgctttaaaaaattctcttaataataaaaaataatagactgctagtcagtaagat 1796
 Db 659 CCAGAGTGCTTTAAAAAATCTCTTAATAAATATAGACTCTGATCAGTAAGATTGAT 718
 QY 1797 attt 1800
 Db 719 ATGT 722

RESULT 14
 BG469506
 LOCUS 602532818F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4660540 5',
 DEFINITION mRNA sequence.
 ACCESSION BG469506
 VERSION BG469506.1 GI:13401781
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 829)

AUTHORS
 TITLE
 JOURNAL
 COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1C41458 row: c column: 05

High quality sequence stop: 795.

FEATURES
 source

1. .829
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4660540"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 186 a 190 c 207 g 246 t
 ORIGIN

Query Match 27.6%; Score 641.2; DB 10; Length 829;
 Best Local Similarity 89.9%; Pred. No. 6.4e-157;
 Matches 732; Conservative 0; Mismatches 78; Indels 4; Gaps 4;

QY 5 gctgcagttctctgattgctgacaggggtggaccagccccctctctctatgtgtgtc 64
 Db 18 GCTGCAGTTCCTCATGGCTGCGCGAGGGTGGACAGCCCGCTTCTTATGTGTGTGTC 77
 QY 65 tctgtcgtgactgtgcttgcagagggcagggagcactgtagtgcctatgagtgagga 124
 Db 78 TACTGCTGACCTGTGCTTTGCCGAGGAGGAAAGCTGCTGTAGTGCCTCATGGTGA 137
 QY 125 gccactggttcacatgagtgctggtggagaaactcattctcagggggcagtgagtg 184
 Db 138 GTCACTGTTCCACATGCGAGTGGTGGTGGAGAACTTATCTCAGGGGGCATGAGGTGG 197
 QY 185 ttgtagtcagtcagaggtgagttggcaactgggaagatcactgaattgcacagtgaa 244
 Db 198 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGAAGATCACTGAATTCACAGTGA 257
 QY 245 ctattcaacttcataacctggagagatctgacccggaggttcaaggcttttgccatg 304
 Db 258 CTCTACTCAACTCGTACACTCTGGAAGATCAGAAACCGGGAATTCATGTTTTCGCC 317
 QY 305 ctcaatggaaagcacaagtacgaagatatattctctataatgggttcatacaatgaca 364
 Db 318 CTCAATGGAAGACACAGGACAAAGTATATTCTCTATTAATGAGTTTCATCCAGTGT 377
 QY 365 tttttgactatttttttcaaatgacggaggttgttttaagacacaaaaattagtagaat 424
 Db 378 TTTCTTGACTTATTTTTCGCAATTCGAGGAGTTGGTTTAATCACCAGAAAATAGTAGA 437
 QY 425 acttaaaaggagagttcttttgatgcagtggttctcagatcttttgataactgtgtgcta 483
 Db 438 ACTTAAAGAGGAGAGTCTTTTGGATGCAGTGTTCGTGGATCCTTTTGTACCTCTGCT 497
 QY 484 attgttgcacaaattattctctccctccctccgtgttcttcgcacggggaatacttgcac 543
 Db 498 ATTGTGCTCTAAATATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 557

```

ACCESSION   AI831562
VERSION     AI831562.1  GI:5452233
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 764)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncilogap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            CGAP Library Preparation: M. Bento Soares, Ph.D.
            DNA Sequencing by: Greg Lennon, Ph.D.
            Clone Sequencing by: Washington University Genome Sequencing Center
            found through the NCI-CGAP clone distribution information can be
            www-bio.illn.gov/bbrp/image/image.html
            Insert Length: 1491 Std Error: 0.00
            Seq primer: --40UP from G1bco
            High quality sequence stop: 454.
FEATURES    source
            Location/Qualifiers
            1..764
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="IMAGE:2405161"
            /clone_lib="NCI-CGAP_Lul9"
            /tissue_type="squamous cell carcinoma, poorly
            differentiated (4 pooled tumors, including primary and
            metastatic)"
            /dev_stage="adult"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
            modified polylinker; 1st strand cDNA was prepared from
            pooled lung tumor tissue, and was then primed with a Not I
            - oligo(dT) primer. Double-stranded cDNA was ligated to
            Eco RI adaptors (Pharmacia), digested with Not I and
            cloned into the Not I and Eco RI sites of the modified
            pT73 vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.
            Fatima Bonaldo."
BASE COUNT  242 a 156 c 133 g 232 t 1 others
ORIGIN
Query Match      28.9%; Score 669.4; DB 9; Length 764;
Best Local Similarity 96.8%; Pred. No. 2.5e-164;
Matches 693; Conservative 0; Mismatches 22; Indels 1; Gaps 1;
QY 1605 caatgagaagtgggtgggaataagtgataaaattttgaaccattccctagtcatttccaa 1664
Db 755 CCCATGAAGAGTGGGTGGGAATAGGCTAAATTTTGAACCATCCCTACGTCATCTTCCAA 696
QY 1665 acttgaacaagaatcagtgtaattcattttatttatttatttatttatttatttatttgcata 1724
Db 695 CTTTGAACAGCATCAAGTGTGTAATTCATCTTATTCTTTATTAAGAAATATCTTTGCATA 636
QY 1725 aattaatcagcccgagtgcttttaaaaaattctcttaaaaaataaataagactcgccta 1784
Db 635 AATTAATCAGCCCGAGAGTGC-TTAAANAATTCTCTTAAANAATAAATAATAGACTCGCTA 577
QY 1785 gtctgaataaagatttgatgatcggtcccccctccggtgcttggatcaggaatgaca 1844
Db 576 GTCAAGTACAAGATTGTAATATGATCGTGCCTCCCGGCTCTTTGATCAGGATGACA 517
QY 1845 tgtgcatcttttcagagacgtgcagacagctggcattctagattacttttacttct 1904
Db 516 TGTGCCATTTTTCAGAGGAGCTGCAGACAGGCTGGCAATCTTAGATTACTTTCTTACTCT 457

```

```

QY 1905 gaaacatggcctgttttggagtcgaggattcaaaagtggtccaccgcgtccctactgc 1964
Db 456 GAAACATGGCCTGTTTGGGAGTCGGGGATTCAAAGGTGGTCCACCGCTGCCCTACTGC 397
QY 1965 aaatggcagtttaattatcttcttcttgcagatggttgcagattgcaatgatccctaac 2024
Db 396 AAATGGCAGTTTAAATCATCTCTTTGGCTTCGAGATGGTTGCAATGATCCCTTAACC 337
QY 2025 aataatggtsagtcctctctctctctctctctctctctctctctctctctctctctct 2084
Db 336 AATAATGGT-AGTCCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 277
QY 2085 aggaagacttctacattagatgtagtggaatgaatgaatgaatgagcttgagtgactga 2144
Db 276 AGGAAGACTTGTACTCTTACAGTGTAGTGAATGAATGAATGAATGAATGAATGAATGAAT 217
QY 2145 gaacagcatatgattctctctctctctctctctctctctctctctctctctctctctct 2204
Db 216 GAACAGCATATGATTTCTTGTCTTTGGGAAAAGAAAGAAATGATCTATGAAATTTGTGGTGG 157
QY 2205 tgtattgagaagataaatacattctctctctctctctctctctctctctctctctctctct 2264
Db 156 TGTATTGASAGATAAATCATCTCTTATGCTCAANTGGAGCTGAATTTGATAAAACCCAA 97
QY 2265 aatacagctatgaatgctggggaagtcttctctctctctctctctctctctctctctct 2320
Db 96 AATACAGCTATGAAGTGGTGGGCAAGTTTACTTTTTTCTGATGTTTCTCTCTCTCTCTCT 41

```

```

RESULT 13
BG424452
LOCUS      6024477/2F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4586368 5',
DEFINITION mRNA sequence.
ACCESSION  BG424452
VERSION    BG424452.1  GI:13330958
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 723)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DTP
            CGAP Library Preparation: Ling Hong/Rubin Laboratory
            CGAP Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM1314 row: p column: 17
            High quality sequence stop: 711.
FEATURES    source
            Location/Qualifiers
            1..723
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="IMAGE:4586368"
            /clone_lib="NIH_MGC_14"
            /tissue_type="renal cell adenocarcinoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT  202 a 160 c 170 g 191 t

```

```

Db 181 GGAGCTAAGGAGCTGGAGTACCCTGAATGTTTGGAAATGACTTCTGAAGATTAGA 240
QY 1281 aaatgctctaaagcagtcataatgacaaagttaacagaggaacacatcagcgccttc 1340
Db 241 AAATGCTCTAAAGCAGTCATCAATGACAAAGATTACAGAGGAACATCATGCGCTCTC 300
QY 1341 caacttcacaggaacgcccgggtggaacgctggaacctggcgtgtctctgggtgaggt 1400
Db 301 CAGCCTTCACAGAGACGCCCGGCTGGAGCCCTGGACCTGGCGTGGTCTGGGTGGAGTT 360
QY 1401 tgtgatgaggcacagggcgccacacactgcccgcagccacgacccctcaacttgta 1460
Db 361 TGTGATGAGCACAGGGCGCCACACTGCGCCCGCAGCCACGACCTCACTTGTA 420
QY 1461 ccagttaccattcctggaagtattggtttctctcttgccgtcgctgacagtgccctt 1520
Db 421 CCAGTACCATTCTTGGAGGTGATGTTCTCTTGGCCGTGCTGGCTGGACGTGCGCTT 480
QY 1521 catcaactttaaatgtgtgcttatggctaccggaatgcttggggaaaaaaggcgagt 1580
Db 481 CATCACTTTAAATGTTGTGCTTATGTGCTACCGGAAATGCTTGGGAAAAAGGCGAGT 540
QY 1581 taagaaacccacaaatccaagacccattgagaagtgggtgggaataaaggtaaatattt 1640
Db 541 TAGAAAGCCCAAAATCCAAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAATTTG 600
QY 1641 gaaccattccct-agtcatttccaaacttgaaacagaatcagtggttaaatcatattat 1699
Db 601 GAACCATTCCTTAAGTCAATTCACAACTTGAAAAACAGAAATCAGTCTCAATTCATCT--A 658
QY 1700 tctattagaataacttgcataataataatcagcccgagctgctttaaataatctc 1759
Db 659 TCTATTACGAAATCTTCGCTACATT-ATCAGGCCAGAGTCTCTT-ACAAATCCCTC 716
QY 1760 ttaataaaaaataagactcgtagtcagtaagatattt 1800
Db 717 TAAATAACAAATAATAGACTCGTAGTCAGAACAGATCTCT 757

```

RESULT 11

```

Bg747143 697 bp mRNA linear EST 15-MAY-2001
LOCUS 602704364F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857831 5',
DEFINITION mRNA sequence.
ACCESSION Bg747143
VERSION Bg747143.1 GI:14057796
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution by: NIH Intramural Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1710 row: o column: 16
High quality sequence stop: 684.
Location/Qualifiers
1..697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4857831"
/clone_lib="NIH_MGC_15"

```

FEATURES

source

```

1..697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4857831"
/clone_lib="NIH_MGC_15"

```

```

/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"

```

```

/notes="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

```

```

BASE COUNT 177 a 168 c 184 g 168 t
ORIGIN

```

```

Query Match 29.3%; Score 679.2; DB 10; Length 697;
Best Local Similarity 99.4%; Pred. No. 6.5e-167;
Matches 692; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

```

QY 910 ggaattgtggtttctctcttgggataaatgctcagaaattccagagagaagaagctatg 969
Db 2 GGAATTGTGTTTCCTTTGGATCAATGGTCTCAGAAATTCAGAGAGAAAGCTATG 61
QY 970 gcaattgctgacttgggcaaatccctcagacagtcctctg99cggtacactggaacc 1029
Db 62 GCAATTGCTGATGCTTGGGCAAAATCCCTCAGACAGTCTCTGTGGCGGTACACTGGAAC 121
QY 1030 cgaccatcgaatcttgcgaacacacgatacttcttaagtggctaccaccaaacgactcg 1089
Db 122 CGACCATCGAATCTTGGCAACACAGCATCTTGTAAAGTGGCTACCCCAAAACGATCTG 181
QY 1090 ctgtgtcaccgatgacccgctgcttctatcaaccatgctggttcccatggtgtttatgaa 1149
Db 182 CTGTGTCACCGATGAGCCGCTTATCACCATGCTGTGTCCTCATGTTTATGAA 241
QY 1150 agcatatgcaatgcgttcccatggatgatgccttgtttgtgtagcatgagtgacaat 1209
Db 242 AGCATATGCAATGGCGTTCCCATGATGATGCCCTTGTGTTGGTGTATCAGATGGACAAT 301
QY 1210 gcaaacgcatgagactaaggagctggagtgacccctgaatgttcttgaatgactctt 1269
Db 302 GCAAAAGCCATGAGACTAGGAGCTGGAGTGACCTGATGTCTGGAATGACTTCT 361
QY 1270 gaagatttagaaatgctctaaagcagtcataatgacaaagtacaaaggaggaacatc 1329
Db 362 GAAGATTTAGAAATGCTTAAAGACGATCATCAATGACAAAAGTTACAAAGGAGAACATC 421
QY 1330 atgcccctctccagccttcaaaagaccccgctgtagcgcctggacccctggcgtgttc 1389
Db 422 ATGCCCTCTCCAGCTTTCACANGACGCCCGCTGGAGCCGCTGGACCTGGCCGCTGTC 481
QY 1390 tgggtgagtttgtgatgaggcacaaggcgccacacactgcccgcagccacgac 1449
Db 482 TGGGTGGAGTTTGTGATGAGCACAAAGGGCGCCACACTGCGCCCGCAGCCACGAG 541
QY 1450 ctcaactgtaccagtcaccattccttgacgtgattggttctctcttggcgtcggtcg 1509
Db 542 CTCACCTGGTACCAGTACCATTCCTTGGACGTGATGGTTTCTCTTGGCCGCTGTCG 601
QY 1510 acagtggccttcacacctttaaatgttggcttatggctaccggaaatgcttggggaaa 1569
Db 602 ACAGTGGGCTTCATCACTTTAAATGTTGTCTTATGCTACCGGAAATGCTTGGGAAA 661
QY 1570 aaagggcagtaagaagcccaaatccaaagacc 1605
Db 662 AAAGGGCAGTTAAGAAAG-CCACAAATCCAGACC 696

```

RESULT 12

AI831562/c

LOCUS

DEFINITION

```

AI831562 764 bp mRNA linear EST 21-DEC-1999
w39b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405161 3'
similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,
MICROSOMAL (HUMAN);, mRNA sequence.

```


QY 683 gtttttcaaaatgccttag--aaatagcctctgaaattctcc-aaacacctgttacgg 739
 Db 660 GTTTTTCAAAATGCCCTAGGAAATAGCCTCTGAAATNCTCCAAACACCTGTACGG 719
 QY 740 agtatgatctacagcacacataaattgtgtgtgtgcaacgacattttttgact 799
 Db 720 AGTATGATCTCTACAGCCACACATCACTCGGCTCGTGGCAAGGACTGTGCTTGACT 779
 QY 800 atcccaaacccggtgatcccaacatgatcttctcattgtgtgtatcaac--tgccatcagg 857
 Db 780 AT-CCAAACCCGTGATGCCACAAATGATCCTCTTGGGGTATCCACCTGCCGCTACGG 838
 QY 858 aaagccgtgtgctatgaatttga 881
 Db 839 AAAGCCGTGCTATGATTTTGA 862

RESULT 8
 AI307269/c 818 bp mRNA linear EST 08-APR-1999
 LOCUS tb24d12.x1 NCI-CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2055287 3'
 DEFINITION similar to db:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,
 MICROSMAL (HUMAN);, mRNA sequence.

ACCESSION AI307269
 VERSION AI307269.1 GI:4002458
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.lnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 453.
 FEATURES
 source Location/Qualifiers
 1..818
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2055287"
 /clone.lib="NCI-CGAP_Kid12"
 /tissue_type="2 pooled tumors (clear cell type)"
 /lab_host="DH108"
 /note="Organ: kidney; Vector: p7F3D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid5 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (cloneIDs 1323912-1325831, 1471368-1472903 and
 1492104-1493255). Subtraction by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 247 a 177 c 134 g 258 t 2 others
 ORIGIN

Query Match 30.0%; Score 695.8; DB g; Length 818;
 Best Local Similarity 95.9%; Pred No. 3e-171;
 Matches 745; Conservative 0; Mismatches 29; Indels 3; Gaps 3;

QY 1547 gctaccggaaatgcttggggaaaaagggcg-agttaagaagagcccaacaaatccaagacc 1605
 Db 818 GCTACCAGAAATGCTGGGGAAAAAGGGCGAATTTTAAGNAAGCCNCNCAATCCCAAGAAC 759
 QY 1606 catgaagagtggtgggaaataaagtaaaatttgaaccattccctagctcatttcaaa 1665
 Db 758 ATTTGAGAGTGGGTGGGAATAAAGGTAAATTTTGAACCAATTCGCCCAAGTCAATTTCCCAA 699
 QY 1666 ctgaaacacagaatca-gtgttaaatcatttcttatttataagaagaataactttgcata 1724
 Db 698 CTGAAAAACAGAAATCAGGGGTAAAAATTCATTTTCTTATTAAAGAAATACTTTGGCATA 639
 QY 1725 aattaatcagcccc-agagtgtctttaaataattctcttaataaaaaataaagactcgct 1783
 Db 638 AATTAAATCAGCCCCCAAGAGTCTTTTAAATAATCTCTTAAATAAATAAATAGACTCGCT 579
 QY 1784 agtcagtaagatatattgaatatgtatcgtgccccctccggtgtctttgtatcagatgac 1843
 Db 578 AGTCAGTAAAGATATTTGAATATGTATGTCGCCCTCTGGTCTCTTTGATCAGATGAC 519
 QY 1844 atgtccatttttcagagagcgtgcagacagcgtggcattcagattcttcttactt 1903
 Db 518 ATGTGCAATTTTTCAGAGGACGTGCAGACAGCGTGGCAATTCCTAGATTACTTTTCTACTC 459
 QY 1904 tgaacatggtcctgtttggggagtgcggtattcaaaagtggtccaccgcgtccctactg 1963
 Db 458 TGAACATGTCCTGTTTGGGAGTGGCGGATTCAAAGGTGGTCCACGGCTGCCCTACTG 399
 QY 1964 caaatggcagtttaattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2023
 Db 398 CAAATGGCAGTTTAAATCTTAATCTTTTGGCTCTGAGATGTTGGCAATTCATCCTTAAC 339
 QY 2024 caataatggtcagtcctcattctcttcttcttcttcttcttcttcttcttcttcttctt 2083
 Db 338 CAATAATGTCAGTCCCTCATCTCTGCTGCTTCTCATAGGTGCCACCTGCTGTTTAAAG 279
 QY 2084 aaggaacgtgttaccctttagagtgtagtgagtgagtgagtgagtgagtgagtgagtgag 2143
 Db 278 AAGGAGACCTTTGACCTTTAGAGTGTAGGTGAGTGAATGAATGAATGAATGAATGAATGA 219
 QY 2144 agaacgcatatgatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2203
 Db 218 AGAACGATATGATTTCTTCTGCTTTGGGAAAAAAGAAATGATGCTATGAAATGGTGGTG 159
 QY 2204 gtgtattgagaagaataatcattgttcttcttcttcttcttcttcttcttcttcttcttctt 2263
 Db 158 GTGTATTTGAGAAATTAATCAATGCTTATGCTCAATGAGCTGAATTTGATAAAAAACCA 99
 QY 2264 aaatacagctatgaagtgtcgggcaagtcttcttcttcttcttcttcttcttcttcttcttctt 2320
 Db 98 AAATACAGCTATGAAGTCTGGGCAAGTTTACTTTTTTTTCTGATGTTTCTTCACTAACT 42

RESULT 9
 BC289777
 LOCUS 602385071F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4514129 5',
 DEFINITION mRNA sequence.
 ACCESSION BC289777
 VERSION BC289777.1 GI:13045907
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1..(bases 1 to 881)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM10407 row: m column: 12

High quality sequence stop: 746.

FEATURES

Location/Qualifiers
1. .882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4516403"
/tissue.type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 259 a 155 c 202 g 266 t
ORIGIN

Query Match 31.1%; Score 722.4; DB 10; Length 882;
Best Local Similarity 94.5%; Pred. No. 3.3e-178;
Matches 792; Conservative 0; Mismatches 41; Indels 5; Gaps 4;
Qy 1447 gacctcaactgtacagtagcattcttggagcgtgattggttctcttgcgcgtg 1506
Db 1 GACCTACCTGTACCAAGTACCAATTCCTTGGAGCTGATTGCTCTCTTGGCCGTCG 60
Qy 1507 ctgacagtgccctcatccctttaaattgttgccttgcctacccggaatcttggg 1566
Db 61 CTGACAGTGGCTTCATCACCCTTTAAATGTTGTGCTTGTACCGGAATGCTTGGG 120
Qy 1567 aaaaagggcgagttaagaaagcccaaaatccaaagaccattggaagtgggtggaaa 1626
Db 121 AAAAAAGGCGAGTTAAGAAAGCCCAAAATCCAAAGACCCTTGAAGTGGTGGGAAA 180
Qy 1627 taaggtaaaatttgacattccctagctcaattccaaacttgaaacagaaatcagtt 1686
Db 181 TAAGGTAAATTTTGAACCATTCCTAGTCAATTTCCAACTTGAACAGNATCAGTGT 240
Qy 1687 aaattcaatttattcttataaggaataactttgcataaattatgcagcccgagagtt 1746
Db 241 AAATTCATTTTATTTATTAAGGAATACATTTTGCATAATTAATCAGCCCGAGAGTGT 300
Qy 1747 ttaaaaattctcttaataaaaaataatagactcgtctagtaagtaagataatttgaat 1806
Db 301 TTAATAAATTTCTTAATAAATAATATAGACTCGCTAGTCAAGATATTTGATAT 360
Qy 1807 gtatcgcccccccggtgtcttcttgatcagatgacatgtccattttcagagacgt 1866
Db 361 GTATCGTCCCGTCCCGGTGTCTTTGATCAGGATGACATGTCCCATTTTTCAGAGACGT 420
Qy 1867 gcagacagctggcattctagattactttcttactctgaacatggcctgttggaggt 1926
Db 421 GCAGACAGCTGGCATCTTAGATTACTTTTCTTACTCTGAACATGGCTGTGGAGT 480
Qy 1927 ggggattcaaggtgtgccccccgctgccccctactgcaaatggcagttttaattatc 1986
Db 481 GCGGATTCAAAGGTGGTCCCAACCGCTGCCCTACTGCAATGGCAGTTTAAATCTTATC 540
Qy 1987 ttttgctctgcagatgggtgcaattgatctcttaacaaatagttcagtcctctatc 2046
Db 541 TTTTGGCTTCGAGATGGTTGCAATTCATTCCTTAACCAATTAATGGTCAGTCTCATC 600
Qy 2047 tgtctgttcaataggtgcacacttgtgt-gtttaagaagggaagcttcttgcacattag 2105
Db 601 TGCTCTGCTTCATAGTGGCCACTTGTGTGGTTTAAAGAAAGGAAGCTTGTACCTTTAG 660
Qy 2106 agtgtaggtg-aaataaatgaatggcttgggtgcaatgagacagcatgattct-t 2163
Db 661 AGTGTAGTGAATAATGAATGGTTGGAGTGCATGAGAACATATGATTCTCTGG 720

Qy 2164 gctttgggaaaaagaatgatgtatgaaattggtgtggtgtatttgagaagataatc 2223
Db 721 GCTTTGGGAAAAGAAATGATCTATGAAATGGGGTGGTGGATTTGGAGAAAATATC 780
Qy 2224 attgcttatgtcgaatggagctgaatttgataaaaccccaaaatacacgctatgaatg 2281
Db 781 ATGCTTTATGTCAAGAGG--TGATTGTATAAAACCAATACAGGTNTTAAAGTG 836
RESULT 6
BI759297 876 bp mRNA linear EST 25-SEP-2001
LOCUS 603042939F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5183433 5',
DEFINITION mRNA sequence.
ACCESSION BI759297
VERSION BI759297.1 GI:15750875
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cygabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11457 row: n column: 10
High quality sequence stop: 735.

FEATURES

Location/Qualifiers
1. .876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5183433"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."
BASE COUNT 226 a 193 c 205 g 252 t
ORIGIN

Query Match 31.1%; Score 720.8; DB 10; Length 876;
Best Local Similarity 94.6%; Pred. No. 8.6e-178;
Matches 823; Conservative 0; Mismatches 37; Indels 10; Gaps 7;
Qy 1 ctcaactcagttctctgatgcttgacaggtggaccagccctctctatgtg 60
Db 7 CTCAGCTCAGTCTCTGATGCTTGCACAGGTGGACGCCCTCTCTATGTGTG 66
Qy 61 tgtctgtctgctgacctgtggtttgcgagcgagggaagctactggtgagccatggtat 120
Db 67 TGTCTGCTGTGACCTGTGGCTTTGCCGAGCGAGCAAGCTACTGGTAGTGGCATGGAT 126
Qy 121 gggagccactgggtccactcaggtcaggtgtgtgagaaactcattctcaggggcatgag 180
Db 127 GGGAGCCCACTGGTTCCACATGAGTGGTGGTGGAGAAACTCATTTCTCAGGGGCAATGAG 186

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 09:38:48 ; Search time 4821.79 Seconds
(without alignments)
6494.048 Million cell updates/sec

Title: US-09-835-082-3
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgttctcacaact 2320

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1631.6	70.3	2450	11 BC011409	BC011409 Homo sapi
2	789.6	34.0	2306	11 BC012716	BC012716 Mus muscu
3	780.2	33.6	2202	11 AK002629	AK002629 Mus muscu
4	741.2	31.9	868	10 BG402908	BG402908 602418710
5	722.4	31.1	882	10 BG291462	BG291462 602387347
6	720.8	31.1	876	10 B1759297	B1759297 603042939
7	707.4	30.5	884	10 B1762347	B1762347 603049004
8	695.8	30.0	818	9 AI307269	AI307269 tb24d12.x
9	695.4	30.0	881	10 BG289777	BG289777 602385071
10	680.6	29.3	1005	10 BE870911	BE870911 601448862
11	679.2	29.3	697	10 BG747143	BG747143 602704364
12	669.4	28.9	764	9 AI831562	AI831562 wj39b01.x
13	644	27.8	723	10 BG424452	BG424452 602447772
14	641.2	27.6	829	10 BG469506	BG469506 602532818
15	640	27.6	833	10 BG400591	BG400591 602464244
16	639.8	27.6	791	9 AI765716	AI765716 w183f06.x
17	636.8	27.4	658	9 AV652040	AV652040 AV652040

18	635.4	27.4	945	10 BG167885	BG167885 602339857
19	625.6	27.0	724	10 BF031444	BF031444 601559244
20	621.8	26.8	688	10 BG286763	BG286763 602382611
21	618.8	26.7	796	10 BG432363	BG432363 602496688
22	604.8	26.1	861	9 AT829268	AT829268 wk58a10.x
23	601	25.9	647	9 AV692287	AV692287 AV692287
24	596	25.7	763	10 BG399019	BG399019 602440878
25	594.2	25.6	663	9 AV646537	AV646537 AV646537
26	592.4	25.5	859	10 BE249993	BE249993 600943171
27	589.8	25.4	640	9 AI635995	AI635995 wai4e12.x
28	588	25.3	683	9 AI480360	AI480360 tm51d11.x
29	584.8	25.2	812	10 BG401258	BG401258 602465359
30	584	25.2	621	9 AW151709	AW151709 xf68b10.x
31	583	25.1	667	9 AW051715	AW051715 wx27g08.x
32	583	25.1	969	10 BG291839	BG291839 602386068
33	572	24.7	636	9 AI767470	AI767470 wh27a05.x
34	567.4	24.5	581	10 BE379923	BE379923 601159412
35	567	24.4	623	10 BG171502	BG171502 602322042
36	557.8	24.0	743	10 BG386668	BG386668 602454928
37	556.4	24.0	741	10 BG433491	BG433491 602497921
38	550.6	23.7	619	9 AW300312	AW300312 xs59f05.x
39	549.8	23.7	851	10 BG401340	BG401340 602465455
40	543.8	23.4	700	10 BF130433	BF130433 601818721
41	542.8	23.4	608	9 AW469219	AW469219 hc79e05.x
42	542.8	23.4	953	10 BG328092	BG328092 602427143
43	537	23.1	935	10 BF234991	BF234991 602027544
44	535.4	23.1	626	9 AW082933	AW082933 xc04f05.x
45	532.2	22.9	666	10 BF027719	BF027719 601764020

ALIGNMENTS

RESULT 1
LOCUS BC011409 2450 bp mRNA linear HTC 30-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3854370, mRNA.
ACCESSION BC011409
VERSION BC011409.1 GI:15030273
KEYWORDS htc.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2450)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalob@bcm.tmc.edu
Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 13 Row: b Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4507814
This clone has the following problem: incomplete processing.
Location/Qualifiers

FEATURES

```

; APPLICANT: BELKNAP, WILLIAM R
; TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE
; TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
; TITLE OF INVENTION: GLYCOLALDEHYDES IN SOLANACEOUS PLANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NANCY J. PARSONS
; STREET: 800 BUCHANAN ST.
; CITY: ALBANY
; STATE: CA
; COUNTRY: USA
; ZIP: 94710
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/797,226
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PARSONS, NANCY J
; REGISTRATION NUMBER: 40,364
; REFERENCE/DOCKET NUMBER: 0011.97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 559-5731
; TELEFAX: (510) 559-5777
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Solanum tuberosum
; STRAIN: cv. Lemhi Russet
; INDIVIDUAL ISOLATE: SGT1750
; DEVELOPMENTAL STAGE: mature, somatic
; TISSUE TYPE: tuber
; IMMEDIATE SOURCE:
; LIBRARY: lambda gtl1 cDNA library
; CLONE: SGT 1750
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1486
; OTHER INFORMATION: product="solanidine
; OTHER INFORMATION: glucosyltransferase"
; PUBLICATION INFORMATION:
; AUTHORS: Moehs, Charles P.
; AUTHORS: Allen, Paul V.
; AUTHORS: Friedman, Mendel
; AUTHORS: Belknap, William R.
; TITLE: Cloning and expression of solanidine
; TITLE: UDP-glucose glucosyltransferase from potato
; JOURNAL: The Plant Journal
; VOLUME: 11
; ISSUE: 2
; DATE: 1997
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 488
;
; US-08-797-226-1

```

```

Query Match 1.9%; Score 44.8; DB 2; Length 1607;
Best Local Similarity 53.4%; Pred. No. 0.0072;
Matches 94; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1068 gtgggtaccccaaacatctgtgtgaccccgatgacccgtgctcttatcacccatgc 1127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1069 gtgggtaccccaaacatctgtgtgaccccgatgacccgtgctcttatcacccatgc 1128

```

```

QY 1128 tgggtccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgacctt 1187
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1129 tgggtactaatggttcttgggaagccatcactcttggcgtgccaatgatacactggccact 1188
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1188 gtttgggtacagatggacaatgcaaaagcgcacatgagagactaaaggagcctggagtga 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1189 ttatgctgatcaattctcaacagagaaggtagtcgaggttaggggattgggaatca 1244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 15

```

PCT-US92-00282-37
; Sequence 37, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RIITER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714527 CUSH
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US92-00282-37

```

```

Query Match 1.9%; Score 43.6; DB 5; Length 51;
Best Local Similarity 92.0%; Pred. No. 0.0021;
Matches 46; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1444 cacagacctacacgttacacagatccctctggacgtgattggttctct 1493
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 CACGACCTCACCTGGTCCAGTACCACCTCTTTGGATGTGATTGTTTCCT 50

```

```

Search completed: August 21, 2002, 11:59:49
Job time: 32837 sec

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; PCT-US92-00282-20

Query Match 5.8%; Score 135; DB 5; Length 513;
Best Local Similarity 100.0%; Pred. No. 4.2e-29;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 873 ggaattgaagccctacatgaatgctctggagacaatggaattggtgtttctcttggg 932
Db 230 GGAATTTGAAGCCCTACATTAAATGCTCTCTGGAGAACATGGAATGCTGCTTTGGG 289

QY 933 atcaatgctctcagaattccagagagaagaagctatggtcaattgctgcttggcga 992
Db 290 ATCAATGCTCTCAGAAATTCAGAGAGAAGCTATGTGCAATTGCTGCTTTGGCAA 349

QY 993 aatccctcagacagt 1007
Db 350 AATCCCTCAGACAGT 364

RESULT 13
US-08-522-421-1
; Sequence 1, Application US/08522421
; Patent No. 5908973
; GENERAL INFORMATION:
; APPLICANT: Abu-Bakar, Umi Kalsom
; APPLICANT: Bariton, Sarah Louise
; APPLICANT: Gallego-Veigas, Pedro Pablo
; APPLICANT: Gray, Julie Elizabeth
; APPLICANT: Grierson, Donald
; APPLICANT: Lowe, Alexandra Louise
; APPLICANT: Picton, Steve
; APPLICANT: Whotton, Lee Colin
; TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS
; TITLE OF INVENTION: DERIVED THEREFROM
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSED: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:

```



```

; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
PCT-US92-00282-18

```

```

Query Match      7.0%; Score 161.8; DB 5; Length 1190;
Best Local Similarity 51.8%; Pred. No. 1.3e-36;
Matches 414; Conservative 0; Mismatches 382; Indels 3; Gaps 2;

```

```

QY 68 tctgacacctgtggttgcgcaggcagggaagctactgtgtagtgcctcatgtagggagcc 127
DB 140 TGTCTGGCGCAGTGTCTCCATCTGCGGAAGATACTGTGTGATCCAGTGTGATGCGAGCC 199
QY 128 actggttccaccatgaggtcggtggtgagaaactcttccagggggcagtgaggtggtg 187
DB 200 ACTGCTGAGCATGCTTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAATAGTTG 259
QY 188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaagctt 247
DB 260 TCCTAGCACCTGACGCTGTGTGTACATCAGACAGCGAGCATTTTACACCTTGAAGAGCT 319
QY 248 attcaacttcatatccctggaggtatctgacccggagtt--caaggcttttggccatgc 305
DB 320 ACCCTGTGCCAATCCAAAGGGAGGAGTGTGAAGAGTCTTTTGTAGTCTCGGGGCATAATG 379
QY 306 tcaatggaagacaagtagtacygaagtatatattctctattaaatgggttccatacaatgacat 365
DB 380 TTTTGTGAGATGATCTTTCCTGCGAGCGTGTGATCAAAACATACAGCAAAATAAAGG 439
QY 366 ttttgact-tattttttcaaatgacggaggttggtttaagagcaaaaaattagtagaat 424
DB 440 ACTCTGTATGCTTGTCTGCTGCTGCTCCACTTACTGCAACAAGAGAGCTCATGGCT 499
QY 425 acttaaaagagagttctttttgatgcaggtttctctogactcttttgataaactgtgcttaa 484
DB 500 CCCTGGCAGAAAGCAGCTTTGATGTTCATGCTGACGGAGCCCTTTCCTTCTGCGAGCCCA 559
QY 485 tgggtgcaaatattctccctccctccgtggttcttgcagggggaataacttggcact 544
DB 560 TCGTGGCCAGTACCTGTCTGCGCCACTGTATTCCTGTGATGCTGCTGCTGCTGCTGCT 619
QY 545 atcttgaaagagtgcaagtgccctgtctctcttctctatgttcccccgaattctcttag 604
DB 620 TGAATATTGAGGGTACCAGTGCAGCCCAACCCATTCCTCTAGCTGCGCCAGGCTCTCTCT 679
QY 605 ggtctcagatgccatgactttcaaggagagatcagaaacacatcatcagcttgaggg 664
DB 680 CTCATTGAGATACATGACCTTCTGCGAGCGGTGAGACATGCTCATTTGCTTTCAC 739
QY 665 aacatttattatgccacogtttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
DB 740 AGAAGTCTTGTGCGAGCTGTGTTATTCCTCCGCTATGCAACGCTTTCCTCAGAAATTCCT 799
QY 725 aacacacctgttcagagtagtatctctacagccacacatcaattgtgttgcaacgg 784
DB 800 AGAGAGAGTGTACTGTCAGGACCTATTGAGCTCTGCAATCTGTGCTCTTTTGAAGTG 859
QY 785 acttgttttggactatcccaaacccctgtagcccaacatgatcttcaattgtgtgtatca 844
DB 860 ACTTTGTGAAGGATGACCTAGGCGCCATCATGCGCCCAATATGTTTGTGTGGGAATCA 919
QY 845 actgcccacagggaaagcc 863
DB 920 ACTGCTCTACCAAAATCC 938

```

```

RESULT 10
PCT-US92-00282-12
; Sequence 12, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.

```

```

; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1197 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
PCT-US92-00282-12

```

```

Query Match      6.9%; Score 161.2; DB 5; Length 1197;
Best Local Similarity 52.6%; Pred. No. 2e-36;
Matches 421; Conservative 0; Mismatches 373; Indels 6; Gaps 3;

```

```

QY 85 gccagggcagggaagctactgtgtagtgcctatggatggaggagccactgg--ttcaaccatga 142
DB 304 GCTGAGAGTGTGAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 363
QY 143 ggtcgggtggtgagaaactctctcaggggcatgaggtggttagtgcagcaggg 202
DB 364 GGGAGGCTTTCGGGAGCTTCATGCCAGAGGCCACCGAGCGGTGCTCCTCACCAGG 423
QY 203 tgagttggcaactgggaagatcactgaattgcacagtgaaagacttattcaacttcata 262
DB 424 TGAATATGCACATCAAAAGAGAGAAATTTTTCACCTGACAGCCTATGCTGTTCATGGA 483
QY 263 cccctggagatctgcacgggaggttcaagccttttggccatgctcaat--ggaagacac 319
DB 484 CCCAGAGAGGAATTTGATCCGCTTACGCTGGCTTACACTCAAGGGTCTTTTGAACAGAAC 543
QY 320 aagtcaggaatataattctctctataatgggttgcataceatgacatttttgacttatttt 379
DB 544 ATCTTCTGAAGAGATATTTAGAAAGTATGGCAATTTATGAACAATGATATCTTTGGCCCTTC 603
QY 380 tttaaatgtcagaggttgttttaagacaaaaaattagtagaataacttaaggagactt 439
DB 604 ATAGTGTGTGTGAGAGTACTGCAATATGAGGCCCTGATCAGGCACCTGATATGTACTT 663
QY 440 cttttgatgcagtgtttctcgtatccttttttgataactgtggcttaattgttgccaaatatt 499
DB 664 CCTTTGTATGTGTTTAAACAGACCCCGTTAACTCTGTGTGGCGGTGCTGTGCTTAAGTACC 723
QY 500 tctctccctccctccgtggtctctcagcagggggaataacttttgcacatcttgaagagtg 559
DB 724 TCTCGATCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 783

```

Qy 858 aaagcgttgccatgg 874
 ||||| |||
Db 957 GAAGCCACTATCTCAGG 973

RESULT 8

RESOL 8
PCT-US92-00282-14

; Sequence 14, Application PC/TUS9200282

GENERAL INFORMATION:

APPLICANT: OWENS, IDA S.

APPLICANT: RITTER, JOSEPH K.

TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION

; TITLE OF INVENTION: THEREIN.

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DARBY & CUSHMAN

STREET: 1615 L STREET, N.W.

; CITY: WASHI

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20036-5601

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

;
;
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/00282

FILING DATE: 19920111

; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: SCOTT, WATSON T.

; REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: 91532-PCT

; TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0941

TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1066 base pairs

```

TYPE: NUCLEIC ACID

; STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

PCT-US92-00282-14

Query Match

Query Match 7.9%; Score 183.6; DB 3; Length 1066;
Best Local Similarity 53.5%; Pred. No. 6.7e-43;

Best local similarity 55.5%, PIED: NO. 6.7e-43;
Matches 429; Conservative 0; Mismatches 369

Qy	81	cttgcgaggcaggaaagctactggtagtcgccatggatgaggccactggttcacat	140
Db	202	TTGGGCTGAGAGTGAAGAGGTGTGGTGTGCCATTCATGCGACCACTGGCTCAGCAT	261
Qy	141	gagtgctggtggtggaagaactcattcagggggcagtgagtggttagtcatgcacga	200
Db	262	CGGGAGGCTTTCGGGAGCTCCATGCCAGAGGCCACCAAGCATGGTGCTCTACCCCAAGA	321
Qy	201	ggtgagttggcaactgggaagatcactgaattgcacagtgagaacttattcaacttcata	260
Db	322	GTTGAATATATGCATCAAGAAGAGAACATTTTTCACGCTGCACAACTATGCATTTTCGTG	381
Qy	261	taccttgaagatctggaccggaggttcgaagcitttgcgcctgctcaat---ggaaagc	317
Db	382	GACCCCAGATGAATTTGATGCCATGTGCTGGGCCACACTCAACTGTACTTTGAAACAGA	441
Qy	318	acaactacgaagtatatattctctataatgggttcatacaatgacatttttgacttatt	377
Db	442	ACATTTTCTCGAAGAAATTTTTCAGAAGTATGCAATGTTCGAACAATATGTTTGGTCTA	501


```
Db 303 GGTGAATATGTACATCAAGAGAGAACTTTTACACCTGACACAGTATGCCATTTCATG 362
Qy 261 taccctgagagatcggaccggaggttcaaggcttttgcctatgtcgaat---ggaagc 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 GACCCAGAGACAAATTTGATCGCTTTTCTGCTGGTTCACACTCAATGGTTCTTTGAAACAGA 422
Qy 318 acaagtcagagatatactctctattatgggttcatacaatgacatttttgacttatt 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 ACATCTCTGATGAATTTCTAGAGAAATGGCAATATGACAAATATGATATGATGATCAT 482
Qy 378 ttttcaaatgagagaggtttgtttaagacaaataattagtagaactacttaagagag 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 ACATAGGCTTCTGTGAGCTACTGCATAATGAGGCCCTGATCAGGACCTGCTATAC 542
Qy 438 tcttttgatcagtgcttctcgtatccttttgataaactgtgcttaattgttgcgaata 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 TTCTTTGATGTGTTCTACAGAGACCCCTTACCTCTGCGGCGGTGCTGGCTAAGTA 602
Qy 498 tttctccctccctccgtgtcttcgcaggggaatactttgccactatcttgaagaag 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 CTGTGCTGATCTCTGCTGTTTCTTGAGGAACATTTCCATGTGATTTAGACTTTAAGGG 662
Qy 558 tgcacagtcctcctcctctctctatgtcccaagaattctcttaggttctcaatgc 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 CACACAGTGTCCAAACCTCTCTCTATATCTCTAGATTTACTAACGACCAATTCAGACCA 722
Qy 618 catgacttcaagagagagtcaggaacacatcatgctacttggaggaacatttattg 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 723 CATGACATTTCTGCAAGAGGTCAAGAACATGCTCTACCCCTGCGCCCTGTCTACCTTTG 782
Qy 678 ccacgttttttcaaaaagccctagaaatagcctctgaaattctccaaacacctgtac 737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 783 CAGCTGTTTCTGCTCTTATGCAAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 842
Qy 738 ggaatgatctctacacacacacatcaatttgggttgcgaacggacttttggta 797
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 843 AGTGTGATCTTGTACAGCATGATCTGCTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 902
Qy 798 ctatcccaaacccgctgctcccaacatcatctcattgtgtgtatcaactgccatcag 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 903 TTACCCAGGCCGATCATGCCCAACATGGTCTTCATTTGGGGGCACTCACTGTGCCAACGG 962
Qy 858 aaagccgttgcctatggaattgaagccatcata 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 963 GAAGCCACTATCTCAGGTCTGTATTGGTGGCTTTA 997
```

RESULT 7

```
; Sequence 16 Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; TITLE OF INVENTION: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESS: CUSHMAN DABBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: YES
; PCT-US92-00282-16

Query Match      8.0%; Score 186.6; DB 5; Length 1448;
Best Local Similarity 53.8%; Pred. No. 1.1e-43;
Matches 429; Conservative 0; Mismatches 364; Indels 4; Gaps 2;

Qy 81 ctttcgcagagcagggaactactgttagtgcctatggatggaggaacactgggttcaact 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 CTGGGCTCAGAGTGGGAGGCTGCTGGGTACCCATTGATGATGACGACCCACCTGGCTCG-CAT 236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 141 gagggtcggtggtgagaaactcattctcaggggcatgaggtggtgttagtctatgcacga 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GCGGGAGGCGCTCGGGAGCTCCATGCCAGAGGCACACAGCGGTGGTCTCACCCACGA 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 201 ggtgagttgcaactgagggaatcaactgaattgcacaggaagacttatcaactcata 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 GGTGAATATGCATCAACAAAGAGAAATTTTACCCCTGACACACCTATGCCATTTCATG 356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 261 taccctgagagatcgtggaccggaggttcaaggcttttgcctatgctcgaat---ggaagc 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 GACCCAGGATGAATTTGATCGCTTGTGCTGGGCCACACTCACTGATGTAATCTTTGAAACAGA 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 318 acaagtcagagatatactctctatataatgggttcatacaatgacatttttgacttatt 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 ACATTTTCTGAGACATTTTCTAAAAGTATGGCAATTTTGAAAAATTCATCTTTGGTCTTT 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 378 ttttcaaatgacagaggtttgtttaagacaaataattagtagaatacttaagagagag 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 TCATAGGTCTTGTGTGAGCTACTGCATAATGAGGCCCTGATCAGGACCTGATGCTATC 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 438 tcttttgatcagtgcttctcgtatccttttgataaactgtgcttaattgttgcgaata 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 TTCTTTCGATGTGGTTTAAATAGACCCAAATTTACCTCTCTGGGGCAGTGTCTGCTAAGTA 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 498 tttctccctccctccgtgttcttcgcaggggaatactttgccaactatctttaggttccagatgc 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 CCTATCAATTCCTGCTGCTGTTTTTTTGGAGGAACATTCATTCATTTAGACTTTAAGGG 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 558 tgcacagtcctcgtcctctctatgctcccaagaattctcttaggttccagatgc 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 CACACAGTGTCCAAATCCTTCTCTATATTCCTAAGTTACTTAACGACCAATTCACACCA 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 618 catgacttcaagagagagtagtacgggaacacatcatctgacttggaggaacatttattg 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 CATGACATTCCTGCAAAAGGGTCAAGAACATGCTCTACCCCTGCGCCCTGTCTACCTTTG 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 678 ccacgttttttcaaaaagccctagaaatagcctctgaaattctccaaacacctgttac 737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 CCATGCTCTTCTGCTCTTATGCAAGCTTGGCTCTGAGCTTTTTCAGAGGGAGGTGTC 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 738 ggaatgatctctacagccacacataatgttgggttggtaacagcagacttttggta 797
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 AGTGTGGATCTTCTCAGCCATGTCATGTGTGGCTGTTTCCAGAGGGGACTTTTGTGATGA 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 798 ctatcccaaacccgctgctcccaacatcatctcattgtgtgtatcaactgccatcag 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 TTACCCAGGCCGATCATGCCCAACATGGTCTTCATTTGGGGGCACTCACTGTGCCAACAG 956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

Db      847  TTACTTCTTACTCTGAACATGCGCTGTTGGAGTCCGGGATCAAGAGTGGTCCCA  906
Qy      1949  cegctgcccctactgcaaatggcagtttaattattatttttggcttctgcagatggttg  2008
Db      907  CGGCTGCCCTACTGCAAAATGGCAGTTTAAATCTTATCTTTTGGCTTCGAGATGGTTG  966
Qy      2009  caatgatcccttaaccaataaagtcagctcctcatctctgctgcttcataagtgccac  2068
Db      967  CAATGATCCTTAACCAATATGGTCAGTCTCATCTCTGCTGCTTCATAGGTGCCAC  1026
Qy      2069  ctgtgtgttttaagaagggaagctttgtacctttagtgtagtgtagtggaatgaatg  2128
Db      1027  CTTGTGCTGTTTAAAGAGGGAAGCTTGTACCTTTAGTGTAGGTGAATGAATGAATG  1086
Qy      2129  gcttgagtgcaactgaacacacacacacacacacacacacacacacacacacacac  2188
Db      1087  GCTTGGATGCTGACGACACACATGATGTTCTGCTTGGGGAAGAAAGATGATGCTA  1146
Qy      2189  tgaatttggttggttggttggttggttggttggttggttggttggttggttggttg  2248
Db      1147  TGAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  1206
Qy      2249  ttgtataaaaccccaataacacacacacacacacacacacacacacacacacacac  2308
Db      1207  TTTGATTAACCCCAANTACAGCTATGAAGTCTGGGCAAGTTTACTTTTTTCTTGAT  1266
Qy      2309  ttctctacaact 2320
Db      1267  TTTCTACAAC 1278

```

RESULT 4

```

US-09-180-852-1
; Sequence 1, Application US/09180852
; Patent No. 6287834
; GENERAL INFORMATION:
; APPLICANT: BELANGER, Alain
; APPLICANT: HUM, Dean W.
; APPLICANT: BEAULIEU, Martin
; APPLICANT: LEVESQUE, Eric
; TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
; FILE REFERENCE: 1259-449
; CURRENT APPLICATION NUMBER: US/09/180,852
; EARLIER FILING DATE: 1999-02-08
; EARLIER FILING DATE: 1997-05-16
; EARLIER FILING DATE: 1997-05-16
; EARLIER FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2107
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(1644)
US-09-180-852-1

```

```

Query Match      14.8%; Score 344.4; DB 4; Length 2107;
Best Local Similarity 53.9%; Pred. No. 5.6e-89;
Matches 811; Conservative 0; Mismatches 671; Indels 24; Gaps 4;

Qy      75  ctgtgtgttgccaggaggaagctactggttagtgccatgtagtgccatgtagtgccatggtt  134
Db      102  ctgtgtgttgccaggaggttggaaggtgctggtgtggtggtggtggtggtggtggtggt  161
Qy      135  caccatgagtcggtggtggaagaaacacacacacacacacacacacacacacacacac  194
Db      162  aaatatgaacacacacacacacacacacacacacacacacacacacacacacacacac  221

```

```

Qy      195  gccagaggtgagttggtgcaactgg-----gaagatcactgcaattgacagtggaagactta  248
Db      222  attctcggctcttattcttgcattgcagtaaaatcactctgtctatttaaatagaagttta  281
Qy      249  ttcaacttcataaccctggagagatciggaccggga-----gttcaaggcttt  296
Db      282  tctacatctttaaactaaaaatgatttgggaagattttttatgaaaaatgtctcgataga  341
Qy      297  tgccactgctcaatggaaagcacaagtacgaagtatatattctctatttaattgggttca  356
Db      342  gacataatgatttcaaaaaatacattttgttcataatttttcaacactacaagaatttg  401
Qy      357  caatgacatttttgactatttttttcaaatgcaggagttttttaaagacaaaaatt  416
Db      402  ttgggaatattctgctataataataaagctctgcaagatgcagttttgacaagaacct  461
Qy      417  agtagaacttaaaagagagttcttttgatgcagtggtttctcgtactcttttgaatactg  476
Db      462  tatgaaaaactacaagagtcataatttgatgctctctggtgagatgcggttaactcctg  521
Qy      477  tggcttaattgtgccaaatatcttccctccctccggtggtcttccgaggggaatact  536
Db      522  ttgtgagctgctggtggaacttaacacacacacacacacacacacacacacacacac  581
Qy      537  ttgcaactatcttg-----aagaaggtgcacagtgccctgctcctcttctctatgtccacg  593
Db      582  ttggtcacacagttgagaagaattggtggagatttctgttccctctctctctatgtacct  641
Qy      594  aatctcttaggttctcagatgccatgaccttcaagagagagagagagagagagagagac  650
Db      642  tgtatgtcgaagaattgaatgacaaatgatttcaaggagggataaaaaatatgata  701
Qy      651  catgacttgaggaacatttattatgccaccggttttttcaaaaaatgccctagaataagc  710
Db      702  tatgtttatttgacttttggtttcaagcatatgctatgaaagtgaggaccagtttta  761
Qy      711  ctctgaaattctcaaacacacctgttaaggagtagtatctctacagccacacacacattg  770
Db      762  tagtgaagttctaggagaccacactacattttgagacacattgggggaagctgaaattg  821
Qy      771  gttgttcgaacgacctttgttttgactatcccaaacccggtgattgccccacatgatctt  830
Db      822  gctcattcgaaacctatgtggattttgaaatttctgcccatttaccacaaattgtatt  881
Qy      831  catgttggttatcaactgccatgccatggaagacggttgctgctatggaattgaaagctac  890
Db      882  tgttgggggacttcaactgtaaacacagccacacccctgctcctaaagaaatggaagatt  941
Qy      891  taatgctctggagaacatggaaattgtgtttctcttcttgggatacaatggctcagaat  950
Db      942  gcagagctctggagaaaattggtattgtgtgtgtttctctggtggttcgtatcagtaac  1001
Qy      951  tccagagaagaagctatggcaattgctgactgtttgggcaaaatccctcaacacagctct  1010
Db      1002  gtcagaagaagtgcacaacatgattgcatcagcccttgcagagatcccaaaaaggttct  1061
Qy      1011  gtggcgttacactggaaaccccgaccatcgaaatttgcgaacacacacacacacacacac  1070
Db      1062  atggagatttgaaggcaagaagcacaactttttaggttcccaatactcgtactgataagtg  1121
Qy      1071  gctcccccaaacgactgctgtgctcaccgactgacccgtgctcttattatccacacgctg  1130
Db      1122  gttaccaccagaaagaccccttggcatcccccaaaaacacacacacacacacacacac  1181
Qy      1131  ttcccatgtgtttatgaaagcatatgcaatggcgttcccatggtgtagatgacctgttt  1190
Db      1182  aaccaaaggcatctagaggcgtactaccatgggataccctatggtgggcatctccctgtt  1241
Qy      1191  tgggtgacagatgagacatgcaaaagcgcagatgagagacaaagagagagagagagag  1250
Db      1242  tgcggataacacatgataacattgtcacaatgaagacaaagggagcagccctcagtg  1301

```


RESULT 15	
ABL01468	
ID	ABL01468 standard; DNA; 874 BP.
XX	
AC	ABL01468;
XX	
DT	15-MAR-2002 (first entry)
XX	
DE	Murine apoptosis related DNA sequence #133.
XX	
XX	Apoptosis; mouse; cancer; autoimmune disease; viral infection;
KK	

Db 361 ggcatcagatggacaatgccaagcgcatggaaactcgggagactggggtagacctgaat 420

XX
DR WPI; 2000-052981/04.
DR P-PSDB; AAY57100.

PT New nucleic acid representing polymorphisms in the human uridine
PT diphosphate glucuronosyltransferase gene, used for diagnosis and
PT evaluation of drug metabolism -

PS Examples; Page 43-44; 63pp; English.

Nucleotide sequences AAZ45110-245118 are exons 1A-1J of the human uridine diphosphate-glucosyltransferase 1 (UGT1) gene. The UGTs are a family of enzymes that catalyse the glucuronic acid conjugation of a wide range of endogenous and exogenous substrates including phenols, alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs result in toxic substances being converted to compounds which are more water soluble and are excreted. The invention relates to and identifies UGT1 polymorphisms (AAZ45004-245041). The polymorphism sequences are useful as probes for detecting UGT1 locus polymorphisms, indicative of altered UGT1 expression or activity. These polymorphisms are associated with Crigler-Najjar and Gilbert syndromes (unconjugated hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene is used to predict the rate of metabolism of UGT1 substrates, possible drug-drug interactions and adverse side effects (i.e. to optimize drug dosage), and to screen for diseases caused by exposure to toxins and to study the effects of polymorphisms on enzymatic activity. The UGT1 sequences, including polymorphisms, can also be used to produce the corresponding protein (or its fragments) or to generate transgenic animals or modified cells e.g. for pharmacogenetic screening.

Sequence 735 BP; 189 A; 179 C; 187 G; 180 T; 0 other;

```
Query Match      31.7%; Score 735; DB 21; Length 735;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	874	gaatttgaagcctacattaatgctctgagaacatggaatttgtgtttctctcttggga	933
Db	1	gaatttgaagcctacattaatgctctgagaacatggaatttgtgtttctctcttggga	60
QY	934	tcaatggtctcgaattccagagagaagaactatgccaattgctgattgcttggcga	993
Db	61	tcaatggtctcgaattccagagagaagaagctatggcaattgctgattgcttggcga	120
QY	994	atccctcagaacagtcgtggtcggttacactggaacccgaccatcgaa tcttgcgacaac	1053
Db	121	atccctcagaacagtcgtggtcggttacactggaacccgaccatcgaa tcttgcgacaac	180
QY	1054	acgatactgttaagtggctaccaccaaaacgatactgcttgggtcaaccgatagccctg	1113
Db	181	acgatactgttaagtggctaccaccaaaacgatactgcttgggtcaaccgatagccctg	240
QY	1114	tttatcaaccatgctggtgttcccaagggtttatgaagaacatgacatggtgcgttccatg	1173
Db	241	tttatcaaccatgctggtgttcccaagggtttatgaagaacatgacatggtgcgttccatg	300
QY	1174	gtgatgatcccttgttgggtgatcaga tggacaatgcaaaacgcatggaagactaaagg	1233
Db	301	gtgatgatcccttgttgggtgatcaga tggacaatgcaaaacgcatggaagactaaagg	360
QY	1234	gctggagtgaacctgaattgtctggaattgactctggaagatttagaaaa tgcctctaaaa	1293
Db	361	gctggagtgaacctgaattgtctggaattgactctggaagatttagaaaa tgcctctaaaa	420
QY	1294	gcagtcataatgacaaaagtctacaaggagaacatcatgcgctctccagccttcacag	1353
Db	421	gcagtcataatgacaaaagtctacaaggagaacatcatgcgctctccagccttcacag	480
QY	1354	gacctcccggtggaacgcctggaacctggtgcgtgcttgcggtggagtttctgtgtagagcaac	1413
Db	481	gacctcccggtggaacgcctggaacctggtgcgtgcttgcggtggagtttctgtgtagagcaac	540

Qy	1414	aaggcgccgacacccgcgcgccccagcccccagcactcaactggtaccagtaacattcc	1473
Db	541	aaggcgccgacacccgcgcgccccagcccccagcactcaactggtaccagtaacattcc	600
Qy	1474	ttgacgagtattggttctctcttggccgtgctgacagtgcccttcacactttaa	1533
Db	601	ttgacgagtattggttctctcttggccgtgctgacagtgcccttcacactttaa	660
Qy	1534	tgttgtgctatggtcaccggaaatgctttggggaaaaaaggcgagtaagaagccac	1593
Db	661	tgttgtgctatggtcaccggaaatgctttggggaaaaaaggcgagtaagaagccac	720
Qy	1594	aaatcccaagaccat	1608
Db	721	aaatcccaagaccat	735
RESULT 14			
ID	AAZ45117		
ID	AAZ45117	standard; DNA; 759 BP.	
AC	AAZ45117;		
CC	AAZ45117;		
DT	28-FEB-2000	(first entry)	
DE	UDP-glucuronosyltransferase 1 (UGT1) exon 1J nucleotide sequence.		
KW	Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe; glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice; unconjugated hyperbilirubinemia; drug metabolism; transgenic animal; pharmacogenetic screening; diagnose; ss.		
KW	Homo sapiens.		
OS	Homo sapiens.		
PN	WO957322-A2.		
PD	11-NOV-1999.		
PF	04-MAY-1999;	99WO-US09702.	
PR	07-MAY-1998;	98US-0084807.	
PA	(AXYS-) AXYS PHARM INC.		
PI	Penny L, Galvin M;		
PI	WPI; 2000-052981/04.		
DR	P-PSDB; AAY57039.		
PT	New nucleic acid representing polymorphisms in the human uridine diphosphate glucuronosyltransferase gene, used for diagnosis and evaluation of drug metabolism		
XX	Examples; Page 42; 63pp; English.		
XX			
CC	Nucleotide sequences AAZ45110-245118 are exons 1A-1J of the human uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a family of enzymes that catalyse the glucuronic acid conjugation of a wide range of endogenous and exogenous substrates including phenols, alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs result in toxic substances being converted to compounds which are more water soluble and are excreted. The invention relates to and identifies UGT1 polymorphisms (AAZ45004-245041). The polymorphism sequences are useful as probes for detecting UGT1 locus polymorphisms, indicative of altered UGT1 expression or activity. These polymorphisms are associated with Crigler-Najjar and Gilbert syndromes (unconjugated hyperbilirubinemia) and drug metabolism. The genotyping of the UGT1 gene is used to predict the rate of metabolism of UGT1 substrates, possible drug-drug interactions and adverse side effects (i.e. to optimize drug dosage), and to screen for diseases caused by exposure to toxins and to study the effects of polymorphisms on enzymatic activity. The UGT1 sequences, including polymorphisms, can also be used to produce the corresponding protein (or its fragments) or to generate transgenic		


```

|||||
Db 601 catatgattcttcttggtgggaaagaaatgatgctatgaattgggtggtgtatt 660
Qy 2211 tgagaagataatcatgtctatgtcaaatgagctgaattgataaaacccaaataca 2270
|||||
Db 661 tgagaagataatcatgtctatgtcaaatgagctgaattgataaaacccaaataca 720
Qy 2271 gctatgaagtgctggcgaagtctactttttttctgtgttctctacaact 2320
|||||
Db 721 gctatgaagtgctggcgaagtctactttttttctgtgttctctacaact 770

RESULT 11
AAZ45116
ID AAZ45116 standard; DNA; 930 BP.
XX
AC AAZ45116;
XX
DT 28-FEB-2000 (first entry)
XX
DE UDP-glucuronosyltransferase 1 (UGT1) exon 1H nucleotide sequence.
XX
KW Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW unconjugated hyperbilirubinemia; drug metabolism; transgenic animal;
KW pharmacogenetic screening; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN W0957322-A2.
XX
PD 11-NOV-1999.
XX
PF 04-MAY-1999; 99NO-US09702.
XX
PR 07-MAY-1998; 98US-0084807.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Penny L, Galvin M;
XX
WPI: 2000-052981/04.
XX
P-PSDB: AAY57098.

New nucleic acid representing polymorphisms in the human uridine
diphosphate glucuronosyltransferase gene, used for diagnosis and
evaluation of drug metabolism
-
Examples; Page 39-41; 63pp; English.

Nucleotide sequences AAZ45110-245118 are exons 1A-1J of the human
uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
family of enzymes that catalyze the glucuronic acid conjugation of a wide
range of endogenous and exogenous substrates including phenols,
alcohols, amines and fatty acids. Many of the reactions catalyzed by UGTs
result in toxic substances being converted to compounds which are more
water soluble and are excreted. The invention relates to and identifies
UGT1 polymorphisms (AAZ45004-245041). The polymorphism sequences are
useful as probes for detecting UGT1 locus polymorphisms, indicative of
altered UGT1 expression or activity. These polymorphisms are associated
with Crigler-Najjar and Gilbert syndromes (unconjugated
hyperbilirubinemia) and drug metabolism. The genotyping of the UGT1 gene
is used to predict the rate of metabolism of UGT1 substrates, possible
drug-drug interactions and adverse side effects (i.e. to optimize drug
dosage), and to screen for diseases caused by exposure to toxins and to
study the effects of polymorphisms on enzymatic activity. The UGT1
sequences, including polymorphisms, can also be used to produce the
corresponding protein (or its fragments) or to generate transgenic
animals or modified cells e.g. for pharmacogenetic screening.

Sequence 930 BP; 228 A; 205 C; 212 G; 285 T; 0 other;

```

```

Query Match 33.0%; Score 766.6; DB 21; Length 930;
Best Local Similarity 91.6%; Pred. No. 2.1e-207;
Matches 811; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 19 atggcttgacacaggggtggacacgccccttctctatgtgtgtctgtcgtgcacctgt 78
Db 1 atggctgacacaggggtggacacgccccttctctatgtgtgtctgtcgtgcacctgt 60
Qy 79 ggtcttgccagagcagggaagcactactgtgtggtggtggtggtggtggtggtggtggt 138
Db 61 ggtcttgctgagcagggaagcgtggtggtggtggtggtggtggtggtggtggtggtggt 120
Qy 139 atgaggtcggtggtggaaacactcattctcagggggcaggtggtggtggtggtggtggtggt 198
Db 121 atgaggtcggtggtggaaacactcattctcagggggcaggtggtggtggtggtggtggtggt 180
Qy 199 gagggtgagtggtggcaactgggaagatcactgaattgcacagtggaagacttattcaacttca 258
Db 181 gagggtgagtggtggcaactgggaagatcactgaattgcacagtggaagacttattcaacttca 240
Qy 259 tataccctgaggtatcagcgggaattcaaggtcttttcccatgctcaatgaaagca 318
Db 241 tataccctgaggtatcagcgggaattcaaggtcttttcccatgctcaatgaaagca 300
Qy 319 caagtacgaagtataattctcttcttcttcttcttcttcttcttcttcttcttcttcttctt 378
Db 301 caagtacgaagtattttctcttcttcttcttcttcttcttcttcttcttcttcttcttctt 360
Qy 379 ttttcaaatgacaggtgtgtttaaagacacacacacacacacacacacacacacacacacac 438
Db 361 ttttgcagtgaggtgtgtttaaagacacacacacacacacacacacacacacacacacac 420
Qy 439 tcttttgatgcagtggttctcgtatcttcttcttcttcttcttcttcttcttcttcttcttct 498
Db 421 tcttttgatgcagtggttctcgtatcttcttcttcttcttcttcttcttcttcttcttcttct 480
Qy 499 tttctccctccctccgtgtgttctgcaggggaatactttgccactatcttgcagaaggt 558
Db 481 tttctccctccctccgtgtgttctgcaggggaatactttgccactatcttgcagaaggt 540
Qy 559 gcacagtgccctcgtcctcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 618
Db 541 gcacagtgccctcgtcctcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 600
Qy 619 atgactttcaagagagagtagcaggaacacacacacacacacacacacacacacacacacac 678
Db 601 atgactttcaagagagagtagcaggaacacacacacacacacacacacacacacacacacac 660
Qy 679 caccgttttttcaaaaatgcctagaaatagcctctgaaattctccaaacacacacacacacac 738
Db 661 cagtatttttccaaaatgcctagaaatagcctctgaaattctccaaacacacacacacacac 720
Qy 739 gagtatgatctctacagcagacacacacacacacacacacacacacacacacacacacacacac 798
Db 721 gcatatgatctctacagcagacacacacacacacacacacacacacacacacacacacacacac 780
Qy 799 tatcccaaaacccgtgagtcacacacacacacacacacacacacacacacacacacacacacac 858
Db 781 tatcccaaaacccgtgagtcacacacacacacacacacacacacacacacacacacacacacac 840
Qy 859 aagcgttgctctatgaaattgaaagccttcaattaatgcttcttctgga 903
Db 841 aagcgttgctctatgaaattgaaagccttcttcttcttcttcttcttcttcttcttcttcttct 885

RESULT 12
AAZ45115
ID AAZ45115 standard; DNA; 951 BP.
XX
AC AAZ45115;
XX
DT 28-FEB-2000 (first entry)
XX

```


Db 541 tagctctgcagatggttgcaattgaccccttaaccacataaattggtcagtcctcatctctgt 600

QY 2050 cctgtcttcattagtgccacctgtgtgttttaaagaagggaagcctttgtaccttttagagt 2109

Db 601 cctgtcttcattagtgccacctgtgtgttttaaagaagggaagcctttgtaccttttagagt 660

QY 2110 tagg-gaaatgaatgaatgcttgagctgcaatgagacacacatgattctctgtcttg 2169

Db 661 tagg-gaaatgaatgaatgcttgagctgcaatgagacacacatgattctctgtcttg 720

QY 2170 gggaaagaatgaatgcttatgaaattggtggtggtgtatttgagaagataatcattgct 2229

Db 721 gggaaagaatgaatgcttatgaaattggtggtggtgtatttgagaagataatcattgct 780

QY 2230 tatg-caaatggagctgaatttgataaaaccccaaaatacacagctataagctctggcga 2289

Db 781 tatg-caaatggagctgaatttgataaaaccccaaaatacacagctataagctctggcga 840

QY 2290 gtttactttttctgatttctctacaact 2320

Db 841 gtttactttttctgatttctctacaact 871

RESULT 9

AAQ33027

ID AAQ33027 standard; DNA; 1667 BP.

XX AC AAQ33027;

XX 27-JAN-1993 (first entry)

XX Exon 5 from the UGT1 gene locus.

DE UGT1A; UGT1BP; UGT1C; UGT1D; UGT1E; UGT1F; UGT1G; isozyme; bilirubin;

KW UDP-glucuronosyl transferase; CN; ss.

XX Homo sapiens.

XX Key

FT misc_RNA

FT 1..20

FT Location/Qualifiers

FT /tag= a

FT /note= "representation of 3.3 kbp of

FT non-sequenced DNA between the sequences

FT represented in AAQ33026 and AAQ33027"

FT exon

FT 291..1402

FT /tag= b

FT /label= Exon_5

FT 291..586

FT misc_RNA

FT /tag= c

FT /note= "encodes part of transferase C-terminal; see CC"

FT polyA_signal

FT 1389..1394

FT /tag= d

XX

PN W09212987-A.

XX

PD 06-AUG-1992.

XX

PF 10-JAN-1992;

XX 92WO-US00282.

XX

PR 10-JAN-1991;

XX 91US-0639453.

XX (USSH) US DEPT HEALTH & HUMAN SERVICE.

XX Owens IS, Ritter JK;

XX

DR WPI; 1992-284593/34.

DR P-PSDB; AAR30166.

XX

PT Isolated gene locus UGT1, DNA segments and diagnostic probes -

PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome

PT types I and II

XX

PS Disclosure; Fig 11; 99pp; English.

XX The isolated gene locus, UGT1, has a sequence of about 10000 bp

CC which represent (1) Exon 1, comprising 6 transcriptional units

CC (UGT1F, E, D, C, BP and A), represented in AAQ27368 and

CC AAQ33020-24 respectively;

CC (2) Exon 2, represented in AAQ33025;

CC (3) Exon 3, represented in AAQ33026;

CC (4) Exon 4, represented in AAQ33026;

CC (5) Exon 5, represented in AAQ33027; and

CC (6) about 69 kb of non-sequenced DNA.

CC Six unique N-termini of 286-289 amino acids are encoded by

CC the six different first exons and identical C-termini of 246 amino

CC acids are encoded by the common exons 2-5. The UGT1 gene locus

CC encodes a family of UDP-glucuronosyl transferase isozymes, two of

CC which metabolise bilirubin.

CC Patients having Crigler-Najjar Syndrome (CN) Type I, have a

CC mutation present in the second common exon.

XX

XX Sequence 1667 BP; 487 A; 316 C; 370 G; 474 T; 20 other:

Query Match 34.4%; Score 798; DB 13; Length 1667;

Best Local Similarity 89.1%; Pred. No. 3.4e-216;

Matches 1003; Conservative 0; Mismatches 20; Indels 103; Gaps 9;

QY 1291 aaagcagtcacatgaacaaagttaacaggagaacatcatgcctctccagcctccac 1350

Db 267 aacgaaactgtctgtgttttagttacaggagaaacatcatgcctctccagcctccac 326

QY 1351 aaggaacgc-cggtaggcgcctggaacctggtccggtgtctctgggtgaggttgatgagg 1410

Db 327 aaggaacgc-cggtaggcgcctggaacctggtccggtgtctctgggtgaggttgatgagg 386

QY 1411 cacaaaggcgcgcacacactgcgccgcagccacacacacacacacacacacacacac 1470

Db 387 cacaaaggcgcgcacacactgcgccgcagccacacacacacacacacacacacacac 446

QY 1471 tcttggaactgattggttctctctggtcgtgct-gacagtgccctccacacct 1529

Db 447 tcttggaactgattggttctctctggtcgtgctggtcgtggtcgtggtcgtggtcgtggt 506

QY 1530 taatgtgtgttattgctaccggaatgtcttggtgggaaagggcgaggttaagaagc 1589

Db 507 taatgtgtgttattgctaccggaatgtcttggtgggaaagggcgaggttaagaagc 566

QY 1590 ccacaaatc-caagaccatctgagaagtgggtgggaaataagtaaaatttgaaccattc 1649

Db 567 ccacaaatc-caagaccatctgagaagtgggtgggaaataagtaaaatttgaaccattc 626

QY 1650 cctagtcatctccaaactgaaacagaaatcagtggttaaatcattttatttttaag 1709

Db 627 cctagtcatctcc-aacttgaaacagaaatcagtggttaaatcattttatttttaag 685

QY 1710 gaaatactttgcataaataatcagcccgagagtgcttttaaaaaattctcttaaaaaa 1769

Db 686 gaaatac-ttgcataaataatcagcccgagagtgcttttaaaaaattctcttaaaaaa 743

QY 1770 ataagactcgtcagtagtaaaagataattgaataatgtatcgtgccccctcgtgtct 1829

Db 744 ataagactcgtcagtagtaaaagataattgaataatgtatcgtgccccctcgtgtct 803

QY 1830 ttgacaggatgacatgtgccatttttcagagagcgtgcagacagcgtggtcattcttag 1889

Db 804 ttgacaggatgacatgtgccatttttcagagagcgtgcagacagcgtggtcattcttag 862

QY 1890 taatttttactctgaaacatggtcgtgttggtggagtcggggattcaaaaggtggtccac 1949

Db 863 taatttttactctgaaacatggtcgtgttggtggagtcggggattcaaaaggtggtccac 922

QY 1950 cgtgccccctactgcgaatgacagttttaaattcttcttttggctctcgcagatggtgc 2009

Db 923 ggct-cctcactgcgaatgacag-tttaaattcttcttttggctctcgcag-tggtgc 979


```

Db 541 gacgtgcagacaggctgctgattctagattacttttcttactgaaacatggcctgttg 600
Qy 1922 ggaagtcggagattcaaggctgcccacgcgtgccctactgcaaatgacagttttaac 1981
Db 601 ggaagtcggagattcaaggctgcccacgcgtgccctactgcaaatgacagttttaac 660
Qy 1982 ttacttttggctctcagatggttgcaattgacatcttaacaaataatggtcagtcctc 2041
Db 661 ttacttttggctctcagatggttgcaattgacatcttaacaaataatggtcagtcctc 720
Qy 2042 atctctgctcctcagatgacacactgtgtgttttaagaaggagcgtttgtacct 2101
Db 721 atctctgctcctcagatgacacactgtgtgttttaagaaggagcgtttgtacct 780
Qy 2102 ttgaagctagtgtaaatgaatgaatgctggtgagctgacactgagacagcatgatcttc 2161
Db 781 ttgaagctagtgtaaatgaatgaatgctggtgagctgacactgagacagcatgatcttc 840
Qy 2162 ttgctttgggaaaaaagaatgatgatgaatggtggtggtgtgtatttgagaagataa 2221
Db 841 ttgctttgggaaaaaagaatgatgatgaatggtggtggtgtgtatttgagaagataa 900
Qy 2222 tcaattcttatgcaaatgagctgaatttgataaaacccaaataacacagctatgaagtg 2281
Db 901 tcaattcttatgcaaatgagctgaatttgataaaacccaaataacacagctatgaagtg 960
Qy 2282 ctgggcaagtttactttttttctgattgttctctacaact 2320
Db 961 ctgggcaagtttactttttttctgattgttctctacaact 999

```

RESULT 8

AAH51404
ID AAH51404 standard; DNA; 1001 BP.

AC AAH51404;

DT 29-AUG-2001 (first entry)

DE Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 295.

Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
microsomal glutathione S-transferase II; malate decarboxylase enzyme;
DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHT; GS;
flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
uridine diphosphate glucuronosyl transferase; UGT2; asthma; hepatotoxicity;
zileuton; ds.

OS Homo sapiens.

PN W0200058508-A2.

PD 05-OCT-2000.

PF 24-MAR-2000; 2000WO-IB00403.

PR 25-MAR-1999; 99US-0126269.

PR 30-APR-1999; 99US-0131961.

PA (GEST) GENSET.

PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;

DR WPI; 2000-638353/61.

PT Polynucleotides comprising sequences from malate decarboxylase

enzyme-related biallelic markers used for genotyping -

PS Claim 13; Page 487-488; 673pp; English.

Sequences AAH51110-AAH51593 represent human DNA fragments which contain biallelic markers. The sequences are related to various human genes including microsomal glutathione S-transferase II (MGSTII), malate decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione reductase/synthase (GSHR/GSHT), flavin-containing monooxygenases (FMO), gamma-glutamyltransferase 5 (G5T5), dipeptidase (DP), glucose 6-phosphate dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine diphosphate glucuronosyl transferases (UGT2). Each of these sequences contains a biallelic marker/polymorphism, which is represented in the sequence as a degenerate/undefined base. The genes to which the biallelic marker containing sequences are related are involved in drug metabolism. Sequences AAH51594 - AAH51598 represent the genomic sequence of the MGSTII gene and four alternative MGSTII cDNA sequences. AAH62905-AAH62906 are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in an example for the amplification of human genomic DNA fragments. The invention includes a method of genotyping comprising determining the identity of a nucleotide at a DME- or MGSTII-related biallelic marker in a biological sample. The method is used to determine the frequency in a population of an allele of a DME- or MGSTII-related biallelic marker and to select an individual for inclusion in a clinical trial of a drug treatment. The method is also used to detect association between allele and phenotype, and to detect association between haplotype and phenotype. The polynucleotides are used, in hybridization assays, sequencing assays or allele specific amplification assays. The method can be used to determine whether an individual suffers or is at risk of developing asthma or is at risk of developing hepatotoxicity on treatment with zileuton.

XX Sequence 1001 BP; 293 A; 171 C; 210 G; 326 T; 1 other;

Query Match 37.5%; Score 870.6; DB 21; Length 1001;

Best Local Similarity 99.9%; Pred. No. 5.6e-237;

Matches 870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1450 ctcaactgtgtaccagttaccattctctgtgacgtgattgtgttctctctgtgcgtgctg 1509

Db 1 ctcaactgtgtaccagttaccattctctgtgacgtgattgtgttctctctgtgcgtgctg 60

Qy 1510 acagtggctctacacccctttaaattgtgtcttattgtctgctacacgaaatgctgggaaa 1569

Db 61 acagtggctctacacccctttaaattgtgtcttattgtctgctacacgaaatgctgggaaa 120

Qy 1570 aaagggcagtgtaagaagccacaaatccaaagaccattgagaagtggtgggaaataa 1629

Db 121 aaagggcagtgtaagaagccacaaatccaaagaccattgagaagtggtgggaaataa 180

Qy 1630 ggtaaaatttgaaccattcccttagcttccaaacttgaacacagatcagtggttaa 1689

Db 181 ggtaaaatttgaaccattcccttagcttccaaacttgaacacagatcagtggttaa 240

Qy 1690 ttcattttattcttattaaggaataactttgtcatataataatcagccacagtgcttta 1749

Db 241 ttcattttattcttattaaggaataactttgtcatataataatcagccacagtgcttta 300

Qy 1750 aaaaattcttataaaaaataatagactcgtagtcagtaaaagatttgaatatgta 1809

Db 301 aaaaattcttataaaaaataatagactcgtagtcagtaaaagatttgaatatgta 360

Qy 1810 tegtgcctccctccggtgtcttcttgatcaggtgacatgtgccatttttcagaggagtgca 1869

Db 361 tegtgcctccctccggtgtcttcttgatcaggtgacatgtgccatttttcagaggagtgca 420

Qy 1870 gacaggtggcattctagattcttttctactctgaacatggcctgtttggagtgcg 1929

Db 421 gacaggtggcattctagattcttttctactctgaacatggcctgtttggagtgcg 480

Qy 1930 ggaatcaagtggtgccaccgctgccctactgcaaatggcagtttttaattctatttt 1989

Db 481 ggaatcaagtggtgccaccgctgccctactgcaaatggcagtttttaattctatttt 540

Qy 1990 tggctctgcagatggttgcaattgactcttaacaaataatggtcagtcctcactctgt 2049

Db 1990 tggctctgcagatggttgcaattgactcttaacaaataatggtcagtcctcactctgt 2049

XX Key Location/Qualifiers
 FH variation replace (2510,C)
 FT /*tag= a
 FT /standard_name= "Single nucleotide polymorphism"
 FT 2616..14887
 FT /*tag= b
 FT /product= "UDP glycosyltransferase 1"
 FT /note= "this CDS is specifically claimed in claim
 FT 26"
 FT variation replace (2756,T)
 FT /*tag= c
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (2826,A)
 FT /*tag= d
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (3155,G)
 FT /*tag= e
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (9508,C)
 FT /*tag= f
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (9511,T)
 FT /*tag= g
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (10091,T)
 FT /*tag= h
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (10094,T)
 FT /*tag= i
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (10095,C)
 FT /*tag= j
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (10140,C)
 FT /*tag= k
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (14423,T)
 FT /*tag= l
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (14713,T)
 FT /*tag= m
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (14776,T)
 FT /*tag= n
 FT /standard_name= "Single nucleotide polymorphism"
 FT variation replace (14971,C)
 FT /*tag= o
 FT /standard_name= "Single nucleotide polymorphism"
 FT
 PN WO200179230-A2.
 XX
 PD 25-OCT-2001.
 XX
 PD 13-APR-2001; 2001WO-US1273.
 XX
 PD 18-APR-2000; 2000US-197514P.
 XX
 PD (GENA-) GENAISSANCE PHARM INC.
 XX
 PI Chew A, Choi JY, Koshy B, Rounds E;
 XX
 DR WPI; 2002-075063/10.
 DR P-PSDB; AAU10940.
 XX
 PT Genotyping a human UDP glycosyltransferase 1 gene of an individual for
 PT determining the haplotype of an individual, involves determining the
 PT identity of a nucleotide pair at specific polymorphic sites for two
 PT copies of the gene
 XX
 PS Claim 21; Fig 1; 81pp; English.
 XX
 XX The invention relates to genotyping a human UDP glycosyltransferase

CC (UGT1A1) gene of an individual, involving determining for the two copies
 CC of the UGT1A1 gene present in the individual, the identity of the
 CC nucleotide pair at one or more polymorphic sites. The new method is
 CC useful for determining whether an individual has a haplotype or haplotype
 CC pairs, given in the specification. It is useful for improving the
 CC efficacy and reliability of several steps in the discovery and
 CC development of drugs for treating diseases associated with UGT1A1
 CC activity, e.g., Gilbert's syndrome and Crigler-Najjar syndrome, to
 CC validate UGT1A1 as a candidate agent for treating a specific condition or
 CC disease predicted to be associated with UGT1A1 activity, and in the
 CC design of clinical trials of candidate drugs for treating a specific
 CC condition or disease predicted to be associated with UGT1A1 activity. The
 CC method is useful to screen for compounds targeting UGT1A1 to treat a
 CC specific condition or disease associated with UGT1A1 activity. A nucleic
 CC acid (I) comprising a polymorphic variant of a reference sequence for
 CC the UGT1A1 gene or cDNA (II) or its fragment is useful in studying the
 CC expression and function of UGT1A1, and in expressing UGT1A1 protein for
 CC use in screening for candidate drugs to treat diseases related to UGT1A1
 CC activity. (I) or (II) is useful for therapeutic purposes. (II) or a
 CC recombinant organism comprising (II) is useful for studying
 CC expression of the UGT1A1 isogenes in vivo, for in vivo screening and
 CC testing of drugs targeted against UGT1A1 protein, and for testing the
 CC efficacy of therapeutic agents and compounds for Gilbert's syndrome and
 CC Crigler-Najjar syndrome, in a biological system. The present sequence
 CC represents the coding sequence of UDP glycosyltransferase 1.
 XX
 SQ Sequence 18887 BP; 5079 A; 4037 C; 4131 G; 5640 T; 0 other;

Query Match 43.3%; Score 1004.4; DB 24; Length 18887;
 Best Local Similarity 98.4%; Pred. No. 2.8e-274;
 Matches 1014; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 QY 1291 aaagcagtcattcaatgacaaaagttacaagagagacacatcgctctcagccttcac 1350
 DB 14567 aacgaactgcttctgttttagttacaagagacacatcgctctcagccttcac 14626
 QY 1351 aaggaccgcgcgtggagccgctggacccctgctgtctggtgagttgtgtgagag 1410
 DB 14627 aaggaccgcgcgtggagccgctggacccctgctgtctggtgagttgtgtgagag 14686
 QY 1411 cacaaaggcgcgcacacacctggcccgccagccacacacacctggtaccattaccat 1470
 DB 14687 caaaggcgcgcacacacctggcccgccagccacacacacctggtaccattaccat 14746
 QY 1471 tccttgacgtgattgttctcttggcctgctgctgacagtgcccttcacacctt 1530
 DB 14747 tccttgacgtgattgttctcttggcctgctgctgacagtgcccttcacacctt 14806
 QY 1531 aatgtgtgtgtatgctacccgaaatgcttggggaagaaagggcgaggttaagaaagcc 1590
 DB 14807 aatgtgtgtgtatgctacccgaaatgcttggggaagaaagggcgaggttaagaaagcc 14866
 QY 1591 cacaaatccaagaccattgagaagtggtggggaataagtaaaattttgaaccattcc 1650
 DB 14867 cacaaatccaagaccattgagaagtggtggggaataagtaaaattttgaaccattcc 14926
 QY 1651 ctatgtatttccaaactgaaacagacagtcattttaaatttcatttatttatttaag 1710
 DB 14927 ctatgtatttccaaactgaaacagacagtcattttaaatttcatttatttatttaag 14986
 QY 1711 aaatactttgcataaataataacagccagagtgcttttaaaaaattcttcttaaaaaa 1770
 DB 14987 aaatactttgcataaataataacagccagagtgcttttaaaaaattcttcttaaaaaa 15046
 QY 1771 taatagctcctagtcagtaagataatttgaatatatgctgcccctccggtgctt 1830
 DB 15047 taatagctcctagtcagtaagataatttgaatatatgctgcccctccggtgctt 15106
 QY 1831 tgatcaggtgacatgtgccatttttcagagagcgtgcagacagcgtgcatttagatt 1890
 DB 15107 tgatcaggtgacatgtgccatttttcagagagcgtgcagacagcgtgcatttagatt 15166

QY	1861	ggacgctgcagacaggctgcattctctagattacttttttactctgaaacatggcgtgttt	1920
Db	1861	ggacgctgcagacaggctgcattctctagattacttttttactctgaaacatggcgtgttt	1920
QY	1921	gggagtgccggattcaaaagtggtcccaacgctgcacctcaetgcacatggcagttttaat	1980
Db	1921	gggagtgccggattcaaaagtggtcccaacgctgcacctcaetgcacatggcagttttaat	1980
QY	1981	cttatcttttggctctgcagatggtgtgcaattgatccttaaccaataatggctcagtcct	2040
Db	1981	cttatcttttggctctgcagatggtgtgcaattgatccttaaccaataatggctcagtcct	2040
QY	2041	catctctgctctgcttcataagtgccaccttggtgttttaaaggagggaagctttgtacc	2100
Db	2041	catctctgctctgcttcataagtgccaccttggtgttttaaaggagggaagctttgtacc	2100
QY	2101	tttagagtgtaggtggaattgaattgaattggcttgagtgacatggaacagcagcatgtatt	2160
Db	2101	tttagagtgtaggtggaattgaattgaattggcttgagtgacatggaacagcagcatgtatt	2160
QY	2161	cttgctttgggaaaaagaatgatcctatgaaatttgggttggtgtatttgagaagata	2220
Db	2161	cttgctttgggaaaaagaatgatcctatgaaatttgggttggtgtatttgagaagata	2220
QY	2221	atcattgcttatgtccaattggagctggaatttgataaaaaacccaataacagctatgaagt	2280
Db	2221	atcattgcttatgtccaattggagctggaatttgataaaaaacccaataacagctatgaagt	2280
QY	2281	gctgggcaagtttacttttttctgatgtttctctacaact	2320
Db	2281	gctgggcaagtttacttttttctgatgtttctctacaact	2320

RESULT	3
AAQ27369	
ID	AAQ27369 standard; cDNA; 2351 BP.
XX	
XX	AAQ27369;
DT	27-JAN-1993 (first entry)
XX	
DE	HUG-Brl.
XX	
KW	Bilirubin; UDP-glucuronosyltransferase; HUGBrl; HUGBR7;
KW	monoglucuronide; diglucuronide; ss.
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FH	16..784
CDS	/tag= a
FT	polyA_signal
FT	2330..2335
FT	/tag= b
FT	/number= 1
FT	polyA_signal
FT	2338..2343
FT	/tag= c
FT	/number= 2
XX	
PN	WO9212987-A.
PPN	
XX	
PD	06-JAN-1992.
XX	
PF	10-JAN-1992; 92MO-US00282.
XX	
PR	10-JAN-1991; 91US-0639453.
XX	
XX	(USSH) US DEPT HEALTH & HUMAN SERVICE.
PI	Owens IS, Ritter JK;
XX	
DR	WPI; 1992-284503/34.
DR	P-PSDB; AAR26153.
XX	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 12:45:22 ; Search time 682.45 Seconds
(without alignments)
5836.679 Million cell updates/sec

Title: US-09-835-082-3

Perfect score: 2320

Sequence: 1 ctacgtcagttctctgat.....tctcgtatgttctctcaact 2320

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2320	100.0	2320	24	ABA05198 Human uridine 5'di
2	2320	100.0	2320	24	ABA05199 Human uridine 5'di
3	1581.4	68.2	2351	13	AAQ27369 HUG-Br1. Homo sap
4	1576.8	68.0	2368	13	AAQ27370 HUG-Br2. Homo sap
5	1551.8	66.9	1655	23	AAQ27370 HUG-Br2. Homo sap
6	1004.4	43.3	18887	24	AAQ27370 HUG-Br2. Homo sap
7	998.6	43.0	1001	21	AAH51403 DNA encoding UDP g
8	870.6	37.5	1001	21	AAH51403 Human UGT1A7 relat
9	798	34.4	1667	13	AAQ33027 Exon 5 from the UG

10	769.6	33.2	1001	21	AAH51405 Human UGT1A7 relat
11	766.6	33.0	930	21	AAZ45116 UDP-glucuronosyltr
12	741.2	31.9	951	21	AAZ45115 UDP-glucuronosyltr
13	735	31.7	735	21	AAZ45118 UDP-glucuronosyltr
14	659.8	28.4	759	21	AAZ45117 UDP-glucuronosyltr
15	578.8	24.9	874	24	ABL01468 Murine apoptosis r
16	364.2	15.7	391	20	AAZ07208 Human lung tumour
17	364.2	15.7	391	21	AAZ07208 Human lung tumour
18	364.2	15.7	391	23	AAZ07208 Human lung tumour
19	359	15.5	1001	21	AAH51406 Human UGT1A7 relat
20	353	15.2	1001	21	AAH51407 Human UGT1A7 relat
21	345.2	14.9	1854	21	AAZ95200 Human UDP-glucuron
22	345	14.9	1766	23	AAZ95200 Human UDP-glucuron
23	344.4	14.8	2107	19	AAV15900 Uridine diphospho-
24	342	14.7	380	21	AAZ98723 Human colon cancer
25	339.8	14.6	2092	21	AAZ98723 Human UDP-glucuron
26	333.2	14.4	1976	21	AAZ95206 Human UDP-glucuron
27	332.8	14.3	1650	21	AAZ95206 Human carbohydate
28	327.2	14.1	2802	23	AAZ95206 Human carbohydate
29	325.6	14.0	2111	23	AAZ95206 Human carbohydate
30	322.4	13.9	1636	24	AAZ95206 Human carbohydate
31	322.4	13.9	1722	22	AAZ95206 Human carbohydate
32	322.4	13.9	2974	22	AAZ95206 Human carbohydate
33	318	13.7	1859	23	AAZ95206 Human carbohydate
34	314.6	13.6	1713	22	AAZ95206 Human carbohydate
35	302.4	13.0	508	21	AAZ95206 Human carbohydate
36	296	12.8	1001	21	AAH51408 Human UGT1A7 relat
37	282	12.2	2025	23	AAZ91253 Human UGT1A7 relat
38	230.4	9.9	366	21	AAA87085 Rat hepatocyte car
39	222.4	9.6	775	13	AAQ33026 Exons 3-4 from the
40	220.8	9.5	273	21	AAZ98749 Rat hepatocyte car
41	203	8.8	2448	23	AAZ98749 Rat hepatocyte car
42	203	8.8	2448	23	AAZ98749 Rat hepatocyte car
43	203	8.8	2448	23	AAZ98749 Rat hepatocyte car
44	203	8.8	2448	23	AAZ98749 Rat hepatocyte car
45	203	8.8	2448	23	AAZ98749 Rat hepatocyte car

ALIGNMENTS

RESULT 1
ABA05198
ID ABA05198 standard; cDNA; 2320 BP.
XX
AC ABA05198;
XX
DF 04-MAR-2002 (first entry)
XX
DE Human uridine 5'diphosphate glucuronyltransferase UGT1A9 cDNA #1.
XX
KW Human; uridine 5'diphosphate glucuronyltransferase enzyme; UGT1A9;
KW flavopiridol; cancer; gastrointestinal disease; parasitic infection;
KW cytotostatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 19..1611
FT /tag= a
FT /product= "UGT1A9"
XX
PN WO200180896-A2.
XX
PD 01-NOV-2001.
XX
PF 12-APR-2001; 2001WO-US12526.
XX
PR 21-APR-2000; 2000US-0553829.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Ratain MJ, Innocenti F, Iyer L;

68x2